



AGNfitter: A BAYESIAN MCMC APPROACH TO FITTING SPECTRAL ENERGY DISTRIBUTIONS OF AGNs

GABRIELA CALISTRO RIVERA^{1,2}, ELISABETA LUSSO³, JOSEPH F. HENNAWI¹, AND DAVID W. HOGG^{1,4,5,6}¹Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany; calistro@strw.leidenuniv.nl²Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands³INAF Osservatorio Astrofisico di Arcetri, I-50125 Florence, Italy⁴Simons Center for Data Analysis, 160 Fifth Avenue, 7th Floor, New York, NY 10010, USA⁵Center for Cosmology and Particle Physics, Department of Physics, New York University, 4 Washington Place, Room 424, New York, NY 10003, USA⁶Center for Data Science, New York University, 726 Broadway, 7th Floor, New York, NY 10003, USA

Received 2016 April 3; revised 2016 September 22; accepted 2016 September 26; published 2016 December 9

ABSTRACT

We present AGNfitter, a publicly available open-source algorithm implementing a fully Bayesian Markov Chain Monte Carlo method to fit the spectral energy distributions (SEDs) of active galactic nuclei (AGNs) from the sub-millimeter to the UV, allowing one to robustly disentangle the physical processes responsible for their emission. AGNfitter makes use of a large library of theoretical, empirical, and semi-empirical models to characterize both the nuclear and host galaxy emission simultaneously. The model consists of four physical emission components: an accretion disk, a torus of AGN heated dust, stellar populations, and cold dust in star-forming regions. AGNfitter determines the posterior distributions of numerous parameters that govern the physics of AGNs with a fully Bayesian treatment of errors and parameter degeneracies, allowing one to infer integrated luminosities, dust attenuation parameters, stellar masses, and star-formation rates. We tested AGNfitter’s performance on real data by fitting the SEDs of a sample of 714 X-ray selected AGNs from the *XMM-COSMOS* survey, spectroscopically classified as Type1 (unobscured) and Type2 (obscured) AGNs by their optical–UV emission lines. We find that two independent model parameters, namely the reddening of the accretion disk and the column density of the dusty torus, are good proxies for AGN obscuration, allowing us to develop a strategy for classifying AGNs as Type1 or Type2, based solely on an SED-fitting analysis. Our classification scheme is in excellent agreement with the spectroscopic classification, giving a completeness fraction of $\sim 86\%$ and $\sim 70\%$, and an efficiency of $\sim 80\%$ and $\sim 77\%$, for Type1 and Type2 AGNs, respectively.

Key words: galaxies: active – galaxies: nuclei – galaxies: statistics – methods: statistical – quasars: general

1. INTRODUCTION

Active galaxies host in their nuclei (AGNs) the most efficient energy sources in the universe: accreting supermassive black holes (SMBHs) that convert significant fractions of the accreted material rest-mass energies into powerful electromagnetic radiation. The evolution and properties of the host galaxies of AGNs are closely connected to the formation and growth of the SMBHs, as supported by both observational evidence (Magorrian et al. 1998; Marconi & Hunt 2003) and cosmological simulations (Springel et al. 2005; Hopkins et al. 2006). The study of the SMBH/host galaxy co-evolution demands a proper characterization of their properties; for instance the intrinsic AGN luminosity, the covering factor of the nuclear obscuring medium, the stellar mass and star-formation rate (SFR) of the host galaxies and the amount of this emission that is reprocessed by dust. These physical parameters are all encoded in the observed spectral energy distributions (SED) of the sources. In order to constrain these physical parameters and to place AGNs into the context of galaxy evolution, it is fundamental to disentangle the contribution of the AGNs from the host galaxy in the observed SED.

The AGN SED covers the full electromagnetic spectrum from radio to X-rays. The most prominent features are the “infrared bump” at $\sim 3\text{--}20\ \mu\text{m}$, and an upturn in the optical–UV, the so-called “big-blue bump” (BBB; Sanders et al. 1989; Elvis et al. 1994; Richards et al. 2006a; Shang et al. 2011; Elvis et al. 2012; Krawczyk et al. 2013). The BBB is thought to represent the emission from the accretion disk surrounding the SMBH, while the mid-infrared bump is due to the presence of

dust that re-radiates a fraction of the optical–UV disk photons at infrared wavelengths. The presence of this body of gas and dust surrounding the accretion disk (dusty torus) is the foundation of the unified model for AGNs (Antonucci 1993; Urry & Padovani 1995), which explains the differences in their spectral characteristics as an effect of viewing angle with respect to a dusty obscuring torus. These spectral differences classify AGNs into unobscured (Type1) and obscured (Type2) due to the presence or absence, respectively, of broad emission lines in their optical spectra. While in Type1 AGNs the observer has a direct view into the ionized gas clouds Doppler-broadened by the SMBH potential (the broad-line region), in Type2 AGNs the line emission of these clouds is completely or partially extinguished depending on the inclination angle respective of the torus, leaving only the narrow line emission to be observed.

An obvious complication in the study of the host galaxy properties in AGNs is that the emission of the central nuclei outshines the galaxy light; therefore, it becomes extremely difficult to derive constraints on the stellar populations. On the other hand, for obscured AGNs, the host galaxy light may be the dominant component in the optical/near-infrared SED, making it challenging to estimate the intrinsic nuclear power.

A common approach to tackle this problem is to fit the SED with different combinations of theoretical models and/or empirical templates for each individual emission component (e.g., da Cunha et al. 2008). The complexity of the models increases as a more accurate description of the underlying physics is needed. In this way, the number of unknown parameters increases as well, including the unavoidable

existence of degeneracies and correlations among them. A good statistical approach for dealing with parameter degeneracies are Bayesian methods, which also allow the user to obtain reliable confidence ranges for parameter estimates. The importance of a Bayesian study for general SED fitting has provoked the development of several Markov Chain Monte Carlo (MCMC)-based algorithms in the last years, principally in SED studies of quiescent galaxies (e.g., Acquaviva et al. 2011; Serra et al. 2011; Pirzkal et al. 2012; Johnson et al. 2013). These algorithms sample the space to infer galaxy parameters, which have been derived from stellar population synthesis models.

Parallel AGN studies at single wavelength regimes have demonstrated that no monochromatic diagnostic can achieve a complete characterization of AGNs (Juneau et al. 2013) and combined multi-wavelength approaches are necessary for a comprehensive exploration of the physics of active galaxies (e.g., χ^2 -minimization code: Lusso et al. 2012, Bayesian-like code: CIGALE, Ciesla et al. 2015). To date, the publicly available MCMC-based SED-fitting codes in the literature do not include any modeling of the broadband AGN emission in a panchromatic approach, but have focused their analysis exclusively on infrared wavelengths (Sajina et al. 2006; Han & Han 2012; Berta et al. 2013; Hernán-Caballero et al. 2015).

We have thus developed AGNfitter: a probabilistic SED-fitting tool based on MCMC sampling. This code is designed to simultaneously disentangle the physical components of both AGNs and host galaxies from optical–UV to sub-millimeter wavelengths and to infer the posterior distribution of the parameters that govern them. Our Bayesian method allows us to robustly perform this inference in order to recover the parameters with a complete description of their uncertainties and degeneracies through the calculation of their probability density functions (PDFs). The MCMC technique allows us to probe the shape of these PDFs, and the correlations among model parameters, giving more information than just the best-fit and the marginalized values for the parameters. The efficiency and speed of the algorithm makes it capable of treating large samples of AGNs and galaxies, enabling statistical studies of nuclear obscuration in the context of AGN classification.

This paper is structured as follows. In Section 2, we describe the construction of the total AGN and host galaxy models used in the code. In Section 3, we will explain the technical details of the code and the MCMC implementation used. For an illustration of the astrophysical capabilities of AGNfitter, we show in Section 4 the results of the application of the code on synthetic data. In Sections 5 and 6, we describe the sample selection of Type1 and Type2 AGNs of the *XMM*-COSMOS survey and show the results of AGNfitter on this sample.

We adopt a concordance flat Λ -cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$ (Komatsu et al. 2009; Planck Collaboration et al. 2014).

2. THE CODE AGNfitter

AGNfitter produces independent samples of the parameter space to fit the observational data, which is constituted by photometric fluxes ranging from the optical–UV to the sub-millimeter. Following Lusso et al. (2013), we consider four independently modeled components, which cover in total a rest-frame frequency range of $\log \nu = 11\text{--}16$ ($\lambda = 0.001\text{--}100 \mu\text{m}$). These models are specifically the

Table 1
Parameter Space in AGNfitter

Component	Parameter ρ_i	Description	Range
BBB emission	BB	BBB normalization	[0, 20]
BBB emission	$E(B - V)_{\text{bbb}}$	BBB reddening	[0, 1.0]
Torus emission	$\log N_{\text{H}}$ [$\log \text{cm}^{-2}$]	torus column density	[21, 25]
Torus emission	TO	torus normalization	[0, 20]
Stellar emission	τ [Gyr]	exponential SFH timescale	[0.1, 3.0]
Stellar emission	$\log \text{age}$ [log yr]	age of the galaxy	$[10^5,$ $\text{age}(z)]$
Stellar emission	GA	galaxy normalization	[0, 20]
Stellar emission	$E(B - V)_{\text{gal}}$	galaxy reddening	[0, 0.5]
Cold dust emission	\log IRlum [$\log L_\odot$]	parametrization of star- burst templates	[7, 15]
Cold dust emission	SB	starburst normalization	[0, 20]

Note. Description of the parameters and their value ranges. The value of $\text{age}(z)$ is the limiting maximal age for a galaxy at redshift z . Given by the age of the universe at that redshift according to the chosen cosmology.

accretion disk emission (BBB), the hot dust surrounding the accretion disk (the so-called torus), the stellar population of the host galaxy, and emission from cold dust in galactic star-forming regions. We make the assumption that the broadband SED can be constructed as a linear combination of these components with their relative contribution depending on the nature of the source. In order to properly mimic the observed fluxes, the emission templates F_λ need to be integrated against the telescope response curves $S(\lambda)$ corresponding to the spectral bands used, following

$$F_S = \frac{\int \lambda F_\lambda S(\lambda) d\lambda}{\int \lambda S(\lambda) d\lambda}, \quad (1)$$

where F_S is the resulting filtered flux. Depending on the photometric data available, the corresponding filter curves can be chosen from the filter library included in AGNfitter, which is a compilation of COSMOS filters published in <http://www.astro.caltech.edu/~capak/filters/index.html>. New filters that are not available in this list can be easily added by the user.

2.1. Accretion Disk Emission (Big Blue Bump)

The most prominent feature in the AGN SED is the Big Blue Bump (BBB), which is produced by the thermal radiation emitted by matter accreting into the central SMBH (Richstone & Schmidt 1980). AGNfitter models this emission using a modified version of the empirical template constructed by Richards et al. (2006b).

This template is a composite spectrum of 259 Type1 QSO SEDs selected from the Sloan Digital Sky Survey (SDSS, York et al. 2000). Given that the mid-infrared region of the SED will be modeled by independent warm dust templates in AGNfitter, we have disregarded the near-infrared regime originally included in the Richards et al. SEDs. We extrapolated the BBB template from $1 \mu\text{m}$ to longer wavelengths assuming $F_\nu \propto \nu^{-2}$ (i.e., Rayleigh–Jeans tail of a blackbody). Our BBB component is modeled by a single template and the only adjustable parameter is a normalization factor called “BB” (see Table 1) determining its amplitude.

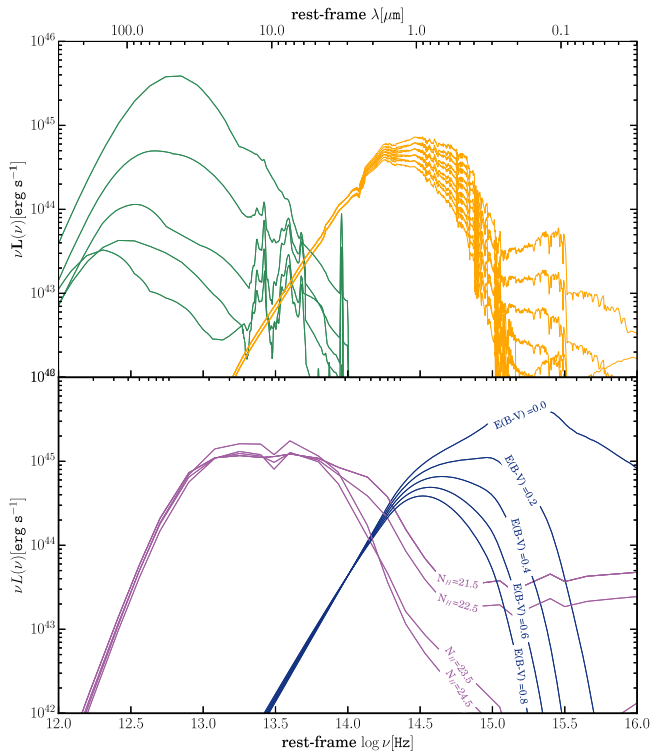


Figure 1. Examples of templates employed in AGNfitter, which models quasar SEDs considering four different components. Upper panel: the green and the orange solid lines correspond to different templates of the starburst and host galaxy components, respectively. Lower panel: the purple and blue solid lines correspond to the hot-dust emission at different column density values $\log N_{\text{H}}$ and the BBB templates with increasing reddening $E(B - V)_{\text{bbb}}$, respectively.

The emitted BBB spectrum can be altered by extinction from dust along the line of sight. We model this effect by applying the Prevot et al. (1984) reddening law for the Small Magellanic Clouds (SMC); which seems to be appropriate for Type-1 AGNs, Hopkins et al. 2004; Salvato et al. 2009) to the template of the emitted BBB SED in our model. Corresponding to the dust screen model, the extinguished flux f_{red} is given by

$$f_{\text{red}}(\lambda) = f_{\text{em}}(\lambda) \times 10^{-0.4A_{\lambda}}, \quad (2)$$

where f_{em} is the emitted flux. The value A_{λ} depends directly on the Prevot reddening law $k(\lambda)$ and the reddening parameter $E(B - V)_{\text{bbb}}$, which is an input parameter in AGNfitter, following

$$A_{\lambda} = k(\lambda)E(B - V)_{\text{bbb}}. \quad (3)$$

The exact function we use for the Prevot’s law is

$$k(\lambda) = 1.39(10^{-4}\lambda)^{-1.2} - 0.38. \quad (4)$$

where λ are in μm .

AGNfitter explores a discrete grid of reddening values in the range of $0 \leq E(B - V)_{\text{bbb}} \leq 1$ in steps of 0.05 and uses Equation (4) to construct the reddening function. Since the MCMC requires a continuous parameter space, we overcome the discrete distribution of the reddening values through a k-Nearest Neighbour Interpolation (kNN), associating any $E(B - V)_{\text{bbb}}$ in the sampling to the nearest $E(B - V)_{\text{bbb-grid}}$ of our templates. A subsample of our BBB templates with different reddening levels is presented in Figure 1 (blue solid lines).

2.2. Hot Dust Emission (torus)

The nuclear hot-dust SED models are taken from Silva et al. (2004). These empirical templates were constructed from a large sample of Seyfert galaxies with reliable observations of the nuclear intrinsic NIR and MIR emission and they are corrected by carefully removing any galaxy contribution. The photometric data obtained from the Seyfert infrared observations were then interpolated based on the radiative transfer models of AGN tori originally presented in Granato & Danese (1994) to obtain full AGN infrared SEDs. This code simulates the emission of dust contents heated by a central source, which has a typical AGN intrinsic spectrum. The infrared SEDs are divided into four intervals of absorption: $N_{\text{H}} < 10^{22} \text{ cm}^{-2}$ for Seyfert 1s (Sy1), $10^{22} < N_{\text{H}} < 10^{23} \text{ cm}^{-2}$, $10^{23} < N_{\text{H}} < 10^{24} \text{ cm}^{-2}$, and $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ for Seyfert 2s (Sy2). As explained in Silva et al. (2004), the main difference between the SEDs as a function of N_{H} is the increment of absorption in the near-IR at $\lambda < 2 \mu\text{m}$ and some mild silicate absorption at $9.7 \mu\text{m}$ for larger N_{H} values. A slight increase of the overall IR emission can also be observed at $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$, as a product of a higher covering factor of the circumnuclear dust at these N_{H} values. In the context of the unified AGN model, the increasing N_{H} values arise from different viewing angles with respect to the torus. We expect thus to observe Type1 AGNs of low N_{H} values with hot dust emission at near-infrared wavelengths, while in the case of Type2 AGNs this near-IR emission may be extinguished by the dust distribution of high N_{H} values. The four templates employed in the code are plotted in Figure 1 with the purple solid line. In order to reduce the degree of discreteness in our parameter space due to the small number of torus templates, we performed a further interpolation between these four templates. We then constructed an empirical function $f(N_{\text{H}})$, which produces a finer parameter grid for the torus models. Our final SED grid consist of 80 templates corresponding to a range of $\log N_{\text{H}} = [21, 25]$ in intervals of $\Delta N_{\text{H}} = 0.05$. Also here, we overcome the discreteness of the N_{H} parameter through a kNN, associating any N_{H} in the sampling to the nearest $N_{\text{H-grid}}$ of our templates. A further parameter adjusting this contribution is the corresponding amplitude parameter “TO” as listed in Table 1.

2.3. Stellar Emission

To construct the library of templates for the stellar emission of the host galaxy, we follow the standard approach and use the stellar population synthesis models by Bruzual & Charlot (2003). These models predict the rest-frame flux of single stellar populations (SSPs) at different ages and for different SFRs. The evolution of the stellar population is computed by assuming a constant metallicity, leaving as free parameters both the age of the galaxy (age) and the star-formation history timescale (τ).

We have assumed a Chabrier (2003) initial mass function and an exponentially declining star-formation history modulated by the timescale τ as $\psi(\text{age}) \propto e^{-\text{age}/\tau}$. For the purposes of this analysis, a set of galaxy templates representative of the entire galaxy population from passive to star-forming is selected. To this end, 10 exponentially decaying star-formation histories (SFHs) with characteristic times ranging from $\tau = 0.1$ to 10 Gyr and a model with constant SFR are included.

For each SFH, a subsample of ages is selected, to avoid both degeneracy among parameters and to speed up the computation. The grid of ages used to calculate the galaxy templates covers ages from 0.2 to 11 Gyr in steps of $\log \text{age} \sim 0.1$. In particular, early-type galaxies, characterized by a small amount of ongoing star formation, are represented by models with values of τ smaller than 1 Gyr and ages larger than 2 Gyr, whereas more actively star-forming galaxies are represented by models with longer values of τ and a wider range of ages from 0.1 to 10 Gyr. An additional constraint on the age implies that for each source, the age has to be smaller than the age of the universe at the redshift of the source.

Altogether, 90 different stellar emission templates from the Bruzual & Charlot (2003) models are included in AGNfitter. A small subsample of these templates is shown in the upper panel of Figure 1 with the orange solid line. Since the MCMC requires a continuous parameter space, we overcome the discrete distribution of the galaxy template parameters through a kNN. In this way, we choose from the grid the nearest combination $(\tau_{\text{grid}}, \text{age}_{\text{grid}})$ corresponding to any pair of parameter values (τ, age) . Additionally, since stellar emission may be altered by extinction along the line of sight, the templates are reddened according to the Milky-Way-like reddening law derived in Calzetti et al. (2000). The reddening values $E(B - V)_{\text{gal}}$ range between 0 and 0.5 with a step of 0.05. Including the extinction effect on the Bruzual & Charlot (2003) SED templates, the total AGNfitter stellar library is composed of 900 stellar templates. Finally, the amplitude of the stellar contribution on the total SED is adjusted by the normalizing parameter ‘‘GA.’’

2.4. Cold Dust Emission from Star-forming Regions

AGNfitter models the emission of cold dust in star-forming regions using the semi-empirical starburst template libraries by Chary & Elbaz (2001) and Dale & Helou (2002). These template libraries represent a wide range of SED shapes and luminosities and are widely used in the literature.

The Chary & Elbaz (2001) template library consists of 105 templates based on the SEDs of four prototypical starburst galaxies (Arp220 (ULIRG); NGC 6090 (LIRG); M82 (starburst); and M51 (normal star-forming galaxy)). They were derived using the Silva et al. (1998) models with the mid-infrared region replaced with ISOCAM observations between 3 and 18 μm (verifying that the observed values of these four galaxies were reproduced by the templates). These templates were then interpolated between the four prototypical starburst galaxies to generate a more diverse set of templates.

The Dale & Helou (2002) templates are updated versions of the widely used Dale et al. (2001) templates. These models involve three components, large dust grains in thermal equilibrium, small grains semistochastically heated, and stochastically heated polycyclic aromatic hydrocarbons (PAHs). They are based on IRAS/ISO observations of 69 normal star-forming galaxies in the wavelength range of 3–100 μm . Dale & Helou (2002) improved upon these models at longer wavelengths using SCUBA observations of 114 galaxies from the Bright Galaxy Sample (see Soifer et al. 1989), 228 galaxies observed with ISOLWS (52–170 μm ; Brauer 2003), and 170 μm observations for 115 galaxies from the ISOPHOT Serendipity Survey (Stickel et al. 2000).

Altogether, the total far-infrared library used in AGNfitter is composed by 169 templates, which are parametrized by their different IR (8–1000 μm) luminosities in the range of $\text{IRLum} = 10^8\text{--}10^{12} L_{\odot}$. The IRLum values were already pre-computed for each template in the original library provided by Chary & Elbaz (2001) and Dale & Helou (2002). It is important to note, that the IRLum parameter is used in AGNfitter solely with the aim of indexing the templates in the library, hence it does not have any physical relevance, and it should not be confused with the computed total infrared luminosity of the sources (denoted as $L_{\text{IR}, 8\text{--}1000 \mu\text{m}}$, see Section 4.4). Also, here we overcome the discrete parametrization of this component through a kNN, associating any IRLum value required by the sampling to the nearest $\text{IRLum}_{\text{grid}}$ of our templates. The amplitudes of this component’s contributions are further adjusted by the normalization parameter ‘‘SB.’’ A small subsample of starburst templates is plotted in Figure 1 with green solid lines.

2.5. Parameter Space

The parameter space describing the total active galaxy emission in AGNfitter is 10-dimensional, as given by the number of parameters listed in Table 1. It is composed by two different types of parameters: (1) physical parameters, which determine the shape of the templates and hence represent non-linear dependencies, and (2) amplitude parameters, on which the model depends linearly and which re-scale the contribution strength of each component. The ranges that we assume each parameter to reside in are listed in Table 1.

3. THE MCMC ALGORITHM

The Bayesian approach in SED-fitting implies that, given the observed photometric data, the posterior probability of the parameters that constitute the model ρ is given by

$$P(\rho|\text{data}) = \frac{P(\text{data}|\rho)P(\rho)}{P(\text{data})}, \quad (5)$$

where ρ is the 10-dimensional vector that includes the parameters listed in Table 1. The function $P(\rho)$ represents our prior knowledge on the distribution of parameters previous to the collection of the data, while the factor $P(\text{data})$ can be regarded as a constant of proportionality, since it is independent of the parameters. The function $P(\text{data}|\rho)$ is the likelihood of the data being observed given the model $\mathcal{L}(\text{data}, \rho) \equiv P(\text{data}|\rho)$.

Provided that the individual photometric measurements are independent and their uncertainties are Gaussian distributed, the likelihood function is given by

$$\mathcal{L}(\text{data}, \rho) \propto \prod_{i=0}^n \exp\left[-\frac{[\text{data}_i - f(\rho)_i]^2}{2\sigma_i^2}\right], \quad (6)$$

where $f(\rho)$ is defined to be the total active galaxy model (linear combination of the four model templates presented in Section 2). One of the main advantages of the Bayesian approach is the possibility of calculating the posterior PDFs of a specific subset of parameters of interest, ignoring other nuisance parameters, in a process referred to as marginalization. The PDF of a single parameter ρ_i is then calculated by integrating over all possible values of the nuisance parameters,

propagating their uncertainties into the final result on $P(\rho_i)$,

$$P(\rho_i) = \int^{k \neq i} d\rho_1 \dots d\rho_k P(\rho|\text{data}). \quad (7)$$

Though the Bayesian approach presents many advantages in theory, AGNfitter requires integrations of a high-dimensional parameter space with parameters related in non-linear functions and it can be extremely computationally intensive in practice. This can be overcome using numerical power, through informed sampling of the parameter space using MCMC algorithms.

Following the Bayesian assumption that unknown parameters can be treated as random variables, the MCMC algorithm produces independent samples of the parameters' posterior PDFs in an efficient way through a random walk in the parameter space. The sequence of visited points in the walk is called a *chain*, where each chain is a Markov process. The ‘‘Markov’’ property ensures that every Monte Carlo (random) step is dependent only on the last visited point having no memory of previous steps.

The random walk is initialized at a position ρ_0 , which we describe in detail in Section 3.2 for the case of AGNfitter. The transition steps between visited points will be decided from a proposal distribution P and the chosen step will be accepted or rejected based on the comparison of the probability of the previous and next trial points. This probability is computed depending on the priors and the likelihood of the model at the current visited point, which results from the calculation of Equation (6). The chosen proposal distribution P is an important factor in the efficiency of the sampling, since it can lead the chain to regions of convergence or to make them diverge toward regions of mainly rejected steps producing heavily correlated samples.

3.1. Emcee

The MCMC core embedded in AGNfitter is the Python package Emcee (Foreman-Mackey et al. 2013), which is a pure-python implementation, slightly modified from the affine-invariant MCMC ensemble sampler developed by Goodman & Weare et al. (2010). In comparison to codes based of the Metropolis–Hastings (MH) algorithm (Metropolis et al. 1953; Hastings 1970), one advantage of Emcee is its affine-invariance property, which implies the clever choice of a proposal distribution, which is invariant under any linear transformation of the posterior distribution to sample. This makes the proposal distribution independent of any possible covariances among the parameters, requiring the hand-tuning of only one or two parameters rather than $\sim N^2$ for a traditional (MH) algorithm in an N -dimensional parameter space. The fact that the efficiency of Emcee is independent of the dimensionality of the model used is a clear advantage for treating a complex problem, such as the one tackled by AGNfitter.

A further property of the Emcee implementation is that the parameter space is explored through a set of chains (*walkers*) that evolve in parallel as an ensemble, rather than through a single chain. This implies that at each step each walker is randomly assigned to a partner walker, moving along lines that connect the single chains to each other. This allows two great features of the code. On one side, this makes the exploration of the parameter space extremely efficient, since the proposal distribution of each current walker is based on the information gained by the rest of the chains. On the other side, Emcee used

the parallel evolution of the chain to allow the distribution of the processes into parallel computing, taking advantage of multi-core processors. Finally, the simple Python interface of Emcee allows more flexibility in its use. A brief discussion comparing several Python statistical packages, which supports our choice of Emcee, can be found in VanderPlas (2014).

3.2. Initialization and Burn in Steps

In AGNfitter, all walkers start in a random position in a region of very small radius around the central values of the parameter ranges (see Table 1). Before running the MCMC sampling, which should return the right PDFs, the walkers need to arrive to the region of interest, around which the maximum likelihood should be located. There is a certain number of steps used for this aim, and the period needed for this initial convergence is called *burn in*. The burn in steps will be neglected for the purposes of sampling the posterior probability distribution of the parameters.

Achieving the convergence of the chains is an important condition for the correct sampling of the target distribution and different methods on judging convergence are discussed in the literature (e.g., Gelman & Rubin 1992; Mary Kathryn Cowles 1996; Lewis & Bridle 2002). As explained in detail in Foreman-Mackey et al. (2013), a good convergence diagnostic consists of measuring the autocorrelation function and more specifically, the integrated autocorrelation time of the chains. The inverse of the autocorrelation time estimates the fraction of the total sample, which has converged into the relevant regions, i.e., the effective fraction, which samples the target distribution. As a feature of Emcee, AGNfitter includes the option of calculating the autocorrelation time using the Python module `acor2`. The longer the autocorrelation time, the more samples are needed during the burn in face to ensure independent samples of the target density. A good convergence test recommended for AGNfitter chains consists of verifying that the autocorrelation time for each parameter is at least a factor of ~ 32 shorter than the chain length.

One recurring issue in parallel evolving chain-ensembles is that some of the chains can get trapped in some uninteresting local, but not global, maxima. In order to deal with this problem, we have introduced multiple burn-in processes. The basic idea is that, after each single burn-in sampling, all walkers are relocated to the point of the highest likelihood (global maximum) visited during the last set, restarting a new sampling set, and avoiding the scenario in which some of the walkers remain in regions of local maxima. This process is iterated ideally until the region of interest (i.e., around the location of the maximum likelihood value) is achieved. Unfortunately, there is no established theory for how long the burn in process should be, and this must instead be determined empirically for the problem at hand. The number of burn-in sets, as well as the number of steps in each burn in, can be set by the user. For the size of the AGNfitter parameter space, we recommend at least two burn-in sets of 10,000 steps each.

3.3. MCMC Steps

After the burn-in period, all further samples are included in the calculations of the posterior PDFs. The length needed for the sampling depends on several factors such as the size and dimension of the parameter space, the quality of the data (number of data points, error sizes), whether the convergence

of the chains has been achieved previously in the burn-in process, and the density of the sampling desired. For the latter, the number of sampling points used for the distribution is not only determined by the number of MCMC steps but also by the number of walkers used for the parameter space exploration.

3.4. Running AGNfitter

We set AGNfitter to use a number of 40 walkers and two burn-in sets each of 10,000 steps as default values. This can be freely changed based on the user’s preferences. After the burn-in phase, the PDFs are sampled with a number of 30,000 steps during the MCMC procedure. For this default configuration, the SED-fitting process takes in total about five minutes per source of computational time in a Macbook Pro with a 2.8 GHz Intel Core i5 processor, using only one core. It should be noted that the time needed for sampling the PDFs is both source and data dependent because the parameter space exploration time depends on how often the proposed steps are accepted or neglected. Since the consideration of each proposed step involves likelihood calculations, the time may increase for sources that are difficult to fit.

For the case of large catalogs, AGNfitter provides the possibility of fitting several sources simultaneously using the multiprocessing Python package, which distributes automatically several fitting processes into a specified number of computer cores. The availability of a multi-core computer would largely increase the efficiency of the code.

3.5. Prior on the Galaxy Luminosity

One of the problems we faced while fitting the optical–UV region of the SED of Type 1 AGNs was the degeneracy between the BBB and galaxies with very bright and young stellar populations. A luminous Type 1 AGN ($L_{\text{bol}} \sim 10^{45-46} \text{ erg s}^{-1}$) can be either modeled with a BBB with zero reddening and negligible galaxy contribution, or with a highly reddened BBB (e.g., $E(B - V) \gtrsim 0.5$) and a star-forming galaxy with significant contribution by the young stellar population. However, such galaxies will have exceptionally high luminosities, much higher than typically observed. Part of this degeneracy is due to the lack of flexibility for the BBB template currently employed: the shape of the BBB (at zero reddening) is fixed, and it does not change with black hole mass.⁷ On the other hand, stellar population synthesis models are much more flexible, since their shape depends on several parameters (e.g., age, τ , metallicity, etc.). To overcome this issue, in AGNfitter, we have implemented a simple prior on the galaxy luminosity based on the galaxy luminosity function. Specifically, we considered the redshift evolution of galaxy Schechter luminosity functions. As estimated by Iovino et al. (2010 see also Kovač et al. 2010), after fitting this redshift evolution, the characteristic absolute magnitude in the B-band of the stellar population M_B^* has the following functional form:

$$M_B^* = -20.3 - 5 \log(h_{70}) - 1.1 \times z. \quad (8)$$

M_B^* includes a redshift evolution of roughly 1 magnitude between $z = 0.1$ and $z = 1$ (see their Section 4 for further details), and it represents the characteristic absolute magnitude of a galaxy at a given redshift. For each fit, we computed the

absolute magnitude from the galaxy template in the B band ($M_{B,\text{fit}}$). This magnitude is then compared with the expected value from Equation (8) at the source redshift. We set the likelihood to zero if $M_{B,\text{fit}}$ is a factor of 10 times brighter than M_B^* . The choice of the threshold factor is rather arbitrary (and it can be changed by the user), but it provides a simple and effective way of preventing possible degeneracies between reddened BBBs and unphysically luminous stellar populations in Type 1 AGNs.

3.6. IGM Absorption

A significant amount of effort has been devoted to determining the AGN SED in the UV (Zheng et al. 1997; Telfer et al. 2002; Scott et al. 2004; Shull et al. 2012). To estimate the *intrinsic* shape of the BBB, one must take into account the intergalactic medium (IGM) absorption by neutral hydrogen along the line of sight (Stevans et al. 2014; Lusso et al. 2015; Tilton et al. 2016). Absorption from intergalactic H I, blueward of Ly α emission in the AGN rest frame, attenuates the source flux both in the Lyman series (creating the so-called H I forest), and in the Lyman continuum at rest $\lambda_r < 912 \text{ \AA}$ (e.g., Moller & Jakobsen 1990). Therefore, we should consider this effect while modeling AGN SEDs. To correct for the intervening Ly α forest and continuum absorption, one possibility is to employ a set of IGM transmission functions (T_λ) calibrated from multiple quasar absorption-line observables (Lusso et al. 2015, see also Stevans et al. 2014; Tilton et al. 2016 for a different approach) over a wide range of redshifts. Such functions can be estimated statistically given the stochasticity of Lyman limit systems (e.g., Worseck & Prochaska 2011; Prochaska et al. 2014) and critically depends on the parametrization of the H I column density distribution (Madau 1995; Meiksin 2006; Inoue et al. 2014; Prochaska et al. 2014). Given that observations of the IGM are not present to constrain T_λ at all redshifts, this implementation is problematic. Although we recognize the importance of correcting broadband photometry for IGM absorption, in order not to largely increase the model’s complexity, we have been rather conservative and simply neglected all rest-frame data at wavelengths⁸ shorter than 1250 \AA (i.e., $\log \nu/\text{Hz} > 15.38$).

4. ANALYSIS OF SYNTHETIC DATA

4.1. Synthetic SEDs

In order to test the response of AGNfitter to AGNs of different obscuration properties, we construct two synthetic SEDs that mimic typical Type1 and Type2 AGNs. These prototypes are built using four model templates of known input parameters. The parameter values are chosen from representative values for the AGN populations and are shown in tables in the left panel of Figure 2 for the Type1 AGNs and of Figure 3 for the Type2 AGNs. In the middle panels of these figures, the four model templates that correspond to those input parameter values are plotted as dashed lines of different colors, where the colors correspond to different physical components similar to Figure 1. The red dashed line is the total model calculated as a linear combination of these four templates. The total model SEDs are integrated against the broadband filter curves that are

⁷ Our BBB template is representative of an optically selected AGN population with an average black hole mass of about $5 \times 10^8 M_\odot$ (Richards et al. 2006a).

⁸ For simplicity, we have considered the effective wavelength for each photometric band, neglecting their widths.

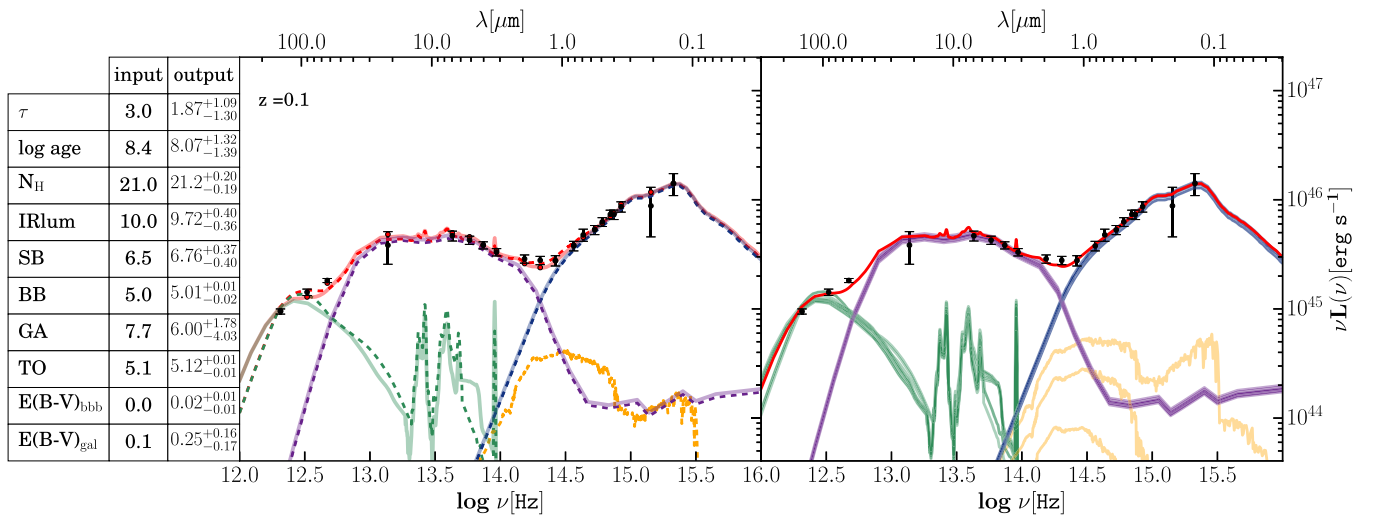


Figure 2. SED output of a synthetic Type1 AGN: the colors of the component SED shapes are assigned similarly to Figure 1, while the linear combination of these, the “total SED,” is depicted as a red line. The table provides the initial input values used for the construction of the mock AGN and the median values of the parameters’ PDFs resulting from the fitting with the respective 16th and 84th percentiles. For the aim of comparison, we plot the SEDs as given by the input parameter values and the output best-fit values in the left panel. The dashed lines are the SEDs corresponding to the input values and the solid transparent lines are the SEDs produced with the output best-fit parameters. In the right panel, we randomly pick eight different realizations from the posterior PDFs and over-plot the corresponding component SEDs in order to visualize the dynamic range of the parameter values included in the PDF. The total SEDs from the realizations are constructed and the mean is depicted as a red line.

included in the COSMOS photometry library to produce mock photometric data points. The uncertainties in the photometry are mimicked through the addition of Gaussian random noise of a standard deviation appropriate from typical uncertainties of each band of the COSMOS photometry (see Section 5). These mock photometric data are plotted in the central and right panels as black dots with error bars.

MOCK Type1 AGNs. The mock Type1 AGN is constructed using the input parameter values specified in the Table of Figure 2. Since the optical–UV regime of the prototypical Type1 SED is dominated by the direct accretion disk emission, we choose a reddening value of $E(B - V)_{\text{bbb}} = 0$, following the observation that Type1 AGNs do not exhibit large amounts of extinction at these wavelengths (Lusso et al. 2012; Merloni et al. 2014). As obscuring media is not abundant in Type1s, the mid-IR/sub-millimeter regime is mimicked by a starburst (cold dust) and a torus template (warm dust) normalized to lower luminosities. We choose these luminosities accordingly to the obscuring fraction for Type1 AGNs in *XMM-COSMOS*, which has been observed to approximately follow the relation $L_{\text{tor}} = 0.5 \times L_{\text{bbb}}$ (Lusso et al. 2013; Merloni et al. 2014). Considering the low optical depth medium observed for Type1 tori (Hatziminaoglou et al. 2008), the torus template is chosen to have the minimum column density of $\log N_{\text{H}} = 21.0$. The galaxy template chosen is determined by parameter values of $\tau = 3$ and $\log \text{age} = 8.4$ (age = 0.251 Gyr). Because of the small stellar contribution to the total emission in Type1 AGNs, the values assumed for the age and τ parameters do not play a significant role in modeling the shape of the total SED. Following these conditions we have modeled the MOCK Type 1 to be a redshift 0.1 bright, unobscured QSO, with direct BBB luminosity $L_{\text{bbb},1-0.1\mu\text{m}} = 2.0 \times 10^{46} \text{ erg s}^{-1}$ and torus luminosity $L_{\text{tor},1-40\mu\text{m}} = 1.0 \times 10^{46} \text{ erg s}^{-1}$, where these luminosities are calculated by integrating over the wavelength ranges specified in the subscripts.

MOCK Type2 AGNs. The synthetic Type2 AGN is constructed following the definition that Type2s lack or have low

accretion disk emission in the optical/UV, since this is obscured on nuclear scales. Therefore, two properties have to be accounted for: (1) The dominant contribution to the total optical/UV luminosity is the stellar emission of the host galaxy, since the AGN contribution (BBB) is obscured. (2) The AGN emission obscured by the hot dust of the torus is reemitted in the IR, and thus the torus emission is expected to represent a large fraction of the accretion disk luminosity in a Type1 AGN (according to the AGN unified model (Antonucci 1993; Urry & Padovani 1995)).

Following condition (1), the BBB template is normalized to a lower luminosity than the Type1 example, producing a total observed luminosity of $L_{\text{bbb-red},1-0.1\mu\text{m}} = 4.4 \times 10^{43} \text{ erg s}^{-1}$ after being extinguished according to a reddening parameter of $E(B - V)_{\text{bbb}} = 0.2$. If we correct for dust extinction, the mock Type 2 AGN presents a dereddened BBB luminosity of $L_{\text{bbb-dered},1-0.1\mu\text{m}} = 1.59 \times 10^{44} \text{ erg s}^{-1}$. The dominant contribution of the stellar emission is chosen to be $L_{\text{gal},1-0.1\mu\text{m}} = 1.9 \times 10^{44} \text{ erg s}^{-1}$ corresponding to $\sim 81\%$ of the total optical/UV contribution. The host galaxy chosen for the mock Type 2 is a galaxy of regular stellar mass $M_{*} = 5.77 \times 10^{10} M_{\odot}$, intermediate age ($\log \text{age} = 8.9$, age = 0.79 Gyr) and follows an exponential SFH with $\tau = 3.0$.

Following condition (2), we model the mock Type 2 torus emission to have the luminosity $L_{\text{tor},1-40\mu\text{m}} = 7.0 \times 10^{43} \text{ erg s}^{-1}$, which corresponds to about 52% of the total MIR luminosity. The chosen template represents a torus of large column density ($\log N_{\text{H}} = 23$), which includes dust self-absorption, meaning that the emission from the inner, hotter part of the torus, emitting at shorter wavelength is highly absorbed (e.g., Silva et al. 1998; Hatziminaoglou et al. 2008).

Following these conditions, we have modeled the mock Type2 AGN as a redshift $z = 2.5$ obscured active galaxy, with an intrinsic luminosity corresponding to $L_{\text{bbb-dered},1-0.1\mu\text{m}} = 1.59 \times 10^{44} \text{ erg s}^{-1}$.

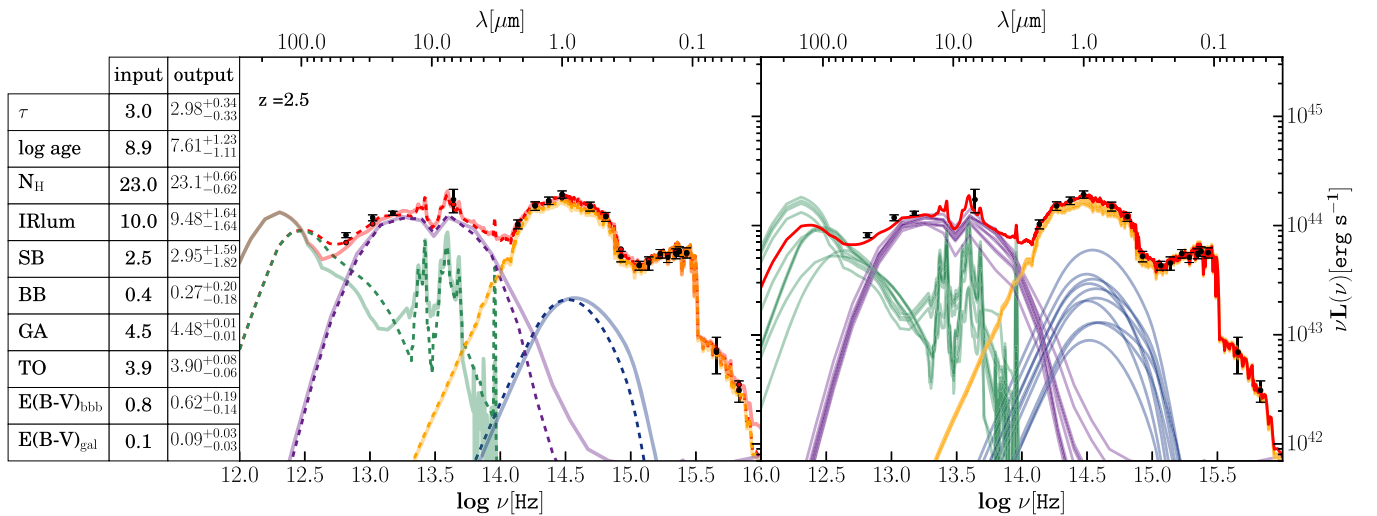


Figure 3. SED output of a synthetic Type2 AGN: the colors of the component SED shapes are assigned similarly to Figure 1, while the linear combination of these, the “total SED,” is depicted as a red line. The table provides the initial input values used for the construction of the mock AGN and the median values of the parameters’ PDFs resulting from the fitting with the respective 16th and 84th percentiles. For the sake of comparison, we plot the SEDs as given by the input parameter values and the output best-fit values in the left panel. The dashed lines are the SEDs corresponding to the input values and the solid transparent lines are the SED produced with the output best-fit parameters. In the right panel, we randomly pick eight different realizations from the posterior PDFs and over-plot the corresponding component SEDs in order to visualize the dynamic range of the parameters values included in the PDF. The total SEDs from the realizations are constructed and the mean is depicted as a red line.

4.2. Recovery of Parameters

After running the code, the comparison between the columns in the tables of Figures 2 and 3 shows a close agreement between the input and output values of the fitting parameters within the uncertainty ranges. This shows clearly the capability of AGNfitter in recovering accurate parameter values as well as error estimates, which encompasses both the noise of the data and the degeneracies among the model parameters.

This can also be seen visually in the middle panels of Figures 2 and 3 where the dashed lines represent the input SEDs used to create the synthetic data points and the solid lines represent the output best-fit SEDs. As shown here, both curves overlap almost completely, showing excellent agreement between the input and output SEDs. Here, we have chosen to plot the best-fit result as the output SED, which is the SED given by the output parameter values of highest likelihood. This result is comparable to the output obtained through other customary SED-fitting tools that use the χ^2 -minimization method. In both examples of Figures 2 and 3 the best-fit solution recovers almost perfectly the input SED shape of each AGN model component. The recovery of the uncertainties is tested through the fitting of 2000 examples of mock Type1 and mock Type2, constructed under the same conditions and input values as the mock AGN in Figures 2 and 3, but adding different random noise to the data according to the photometric errors. Taking into account the Poisson error given by our sample size ($1/\sqrt{N} \sim 2.23\%$), a proper recovery of the uncertainties implies that 95%(68%) of the fitted sources correctly include the “true” values within the 2σ (1σ) confidence regions of the respective posterior PDFs.

For the sample of mock Type1s, the “true” value is recovered within the 2σ confidence region of the parameters for 99.7% of the sample on average. This condition is fulfilled for all parameters without exception. The 1σ confidence regions include the “true” values in 85% of the cases on average. Looking at the parameters separately, we notice that only the τ parameter’s PDFs do not fulfill this condition but in

16% of the cases. Nonetheless, the τ parameter’s 2σ confidence regions include the true value in 100% of the cases.

For the sample of mock Type2s the results are similar. On average, 98.6% of the total sample includes the “true” value within the 2σ confidence regions for all parameters without exception. The 1σ condition is fulfilled on average above the expectations, where 87.7% of the total sample includes the “true” values within this confidence region. Looking at the parameters separately, the age parameter follows this condition in only 20% of the cases, although the “true” value is then recovered within the 2σ confidence region in 100% of the cases.

We conclude that the uncertainties inferred by AGNfitter have been tested to be reliable, though some disagreement with the expectations has been found, which is probably due to the rectangular shape of some parameters’ PDFs, as, for example, the τ and age parameters. Such disagreement is likely to be the consequence of the discreteness level of some model libraries as the galaxy model library used in the code at the time of the publication. In order to improve this, better sampled libraries will be soon available in the online version of the code.

4.3. Posterior PDFs

One of the primary quality assessment products of the algorithm are the PDFs of the fitting parameters. These are summarized in the PDF-triangle of each source and contain a wealth of information on the parameter dependencies.

In Figure 4, the PDF triangle produced from the fitting of the synthetic Type1 AGN is presented as an example. One-dimensional and two-dimensional PDFs of all parameters are depicted in the border and interior parts of the diagram, respectively. The median value (50th percentile) and 16th and 84th percentiles of the PDF are reported by the dashed black lines. The two-dimensional PDFs are shown as probability surfaces, with contours corresponding to 38th, 68th, 88th, and 95th percentiles (equivalent to 0.5, 1, 1.5, and 2σ for Gaussian distributions).

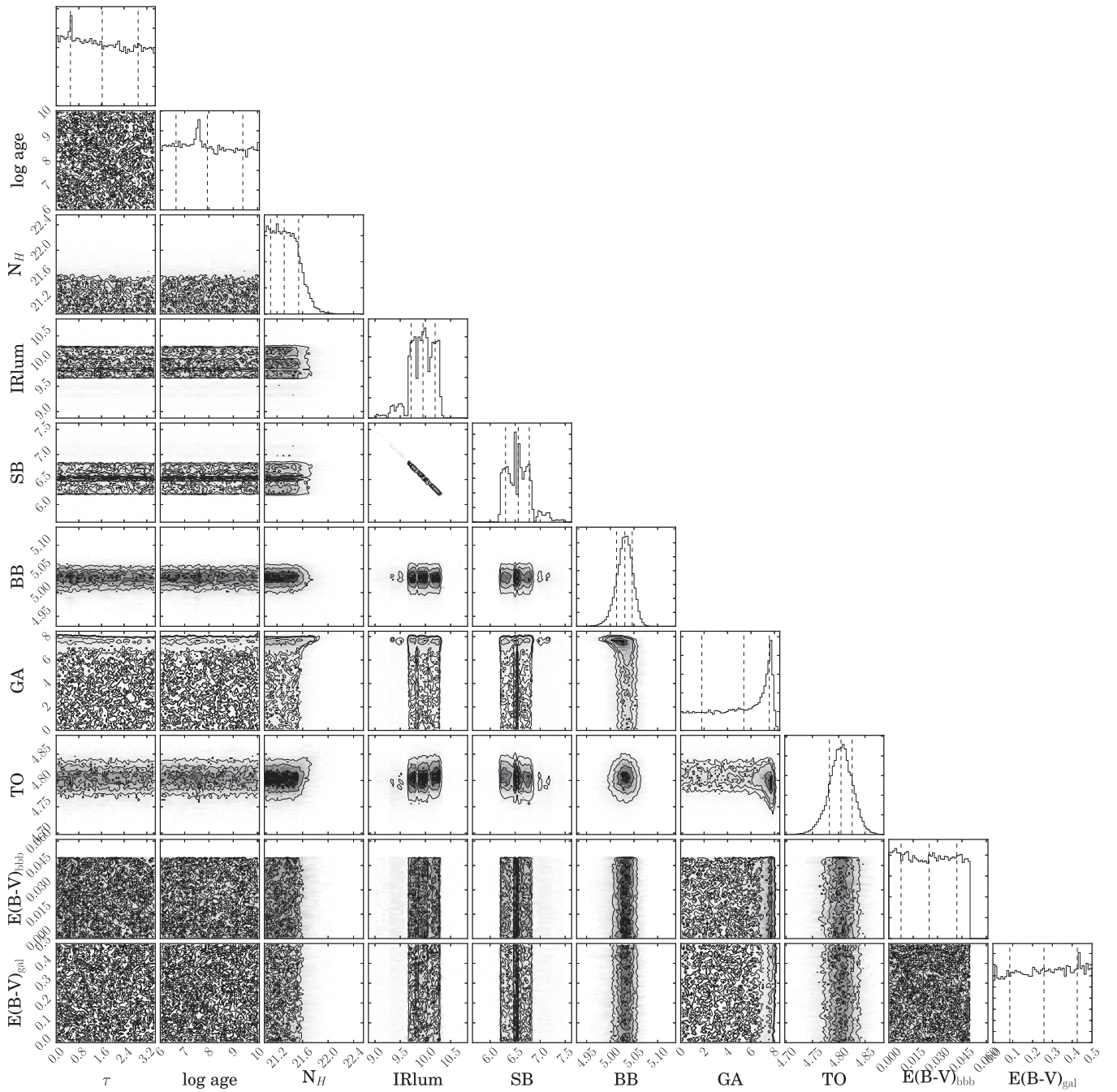


Figure 4. PDF triangle of the synthetic Type 1 AGN shown in Figure 2: one- and two-dimensional PDFs of the components of the parameter space explored by AGNfitter. The outer part of the triangle shows the one-dimensional PDFs for each parameter as histograms. The lines indicate the medium and 1σ uncertainties. The inner part of the triangle shows the two-dimensional PDFs for pairs of parameters as contours. The different contour shades represent 0.5, 1, 1.5, and 2σ .

The shapes of the 1D and 2D PDFs are good indicators of the quality of the inference of a given parameter and hints to the existence of covariances among parameters. As can be noticed for the mock Type1 AGN in Figure 2, parameters associated with the galaxy contribution (τ , age, GA, $E(B - V)_{\text{gal}}$) have very noisy and nearly flat PDFs. This is a consequence of the lack of recognizable galaxy features in the total SED of a Type1 AGN (total SED: red line in 2) and thus, the stellar properties cannot be very accurately constrained. Nonetheless an upper limit for the normalization parameter of this template (parameter GA) is well defined and a maximum contribution to the total luminosity can be inferred. Parameters determining the AGN contributions ($\log N_H$, BB, TO, $E(B - V)$) on the contrary

show more defined posterior distributions and the resulting parameter values, as given by the percentiles, present smaller uncertainties.

As mentioned before, we interpolate between discrete model templates using nearest-neighbor interpolations in order to have a continuous parameter space. This also has an effect on the one-dimensional PDFs of discrete parameters (in this example, IRLum and $E(B - V)_{\text{bb}}$ in Figure 4), which sometimes show abrupt cut-offs on the regions outside the nearest neighborhoods of the discrete values chosen (in this example: IRLum = 10 and $E(B - V)_{\text{bb}} = 0$).

The two-dimensional PDF allows the recognition of covariances of the parameters. This can be clearly observed

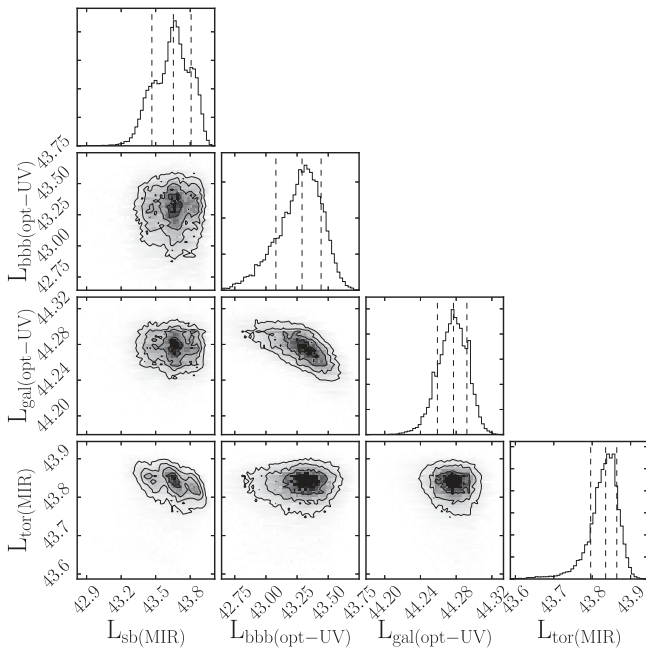


Figure 5. PDF triangle of the synthetic Type2 AGN shown in Figure 3: one- and two-dimensional PDFs of the default AGNfitter output on integrated luminosities is presented. The outer part of the triangle shows the one-dimensional PDFs for each parameter as histograms. The lines indicate the medium and 1σ uncertainties. The inner part of the triangle shows the two-dimensional PDFs for pairs of parameters as contours. The different contour shades represent 0.5, 1, 1.5, and 2σ . The 2D relative posterior of the integrated luminosities $L_{\text{bbb},0.1-1\mu\text{m}}$ and $L_{\text{gal},0.1-1\mu\text{m}}$ show slightly twisted banana shaped contours, revealing that the estimation of the two parameters are degenerate.

for instance in Figure 4 for the parameters IRLum and SB. This strong covariance between these parameters occurs by construction, since as explained in Section 2.4 both the parameters SB and IRLum are normalization parameters of the same model. While SB is a normalization parameter introduced in AGNfitter, IRLum was predefined by the authors of the models to parametrize the SED shapes of the cold dust library based on their total luminosities.

Similarly, complex shaped 2D posterior distributions can result from non-linear dependencies on model parameters. As an example, Figure 5 presents the PDF triangle of the integrated luminosities of the synthetic Type2 AGN. The 2D posterior of the integrated luminosities $L_{\text{bbb},0.1-1\mu\text{m}}$ and $L_{\text{gal},0.1-1\mu\text{m}}$ show slightly twisted banana shaped contours, revealing that the estimation of the two parameters are degenerate. This can be easily understood observing the physical components from which these luminosities are computed in the right panel of Figure 3. Even though the galaxy template dominates the SED, both the reddened BBB and the galaxy template peak around $1\mu\text{m}$, such that both can contribute to the observed SED around this wavelength. This naturally results in a slight degeneracy in their respective integrated luminosities. See Section 4.6 for a more detailed description of the SED plot.

One of the main advantages in estimating the parameters' PDFs is that robust uncertainties on the parameter inference can be now calculated. As describe above, AGNfitter estimates credible intervals from the PDFs (50th, 16th, and 84th percentiles). Following our Bayesian approach, we quote only these uncertainty values from now on in this paper. If requested by the user, AGNfitter also returns the best-fit values for the

parameters, which should be the one with the highest likelihood computed in the chain in the case of convergence. Nonetheless, if the posterior distribution inferred by the sampling is not symmetric, the best-fit value will differ from the median value.

All in all, the true values used for the construction of the mock Type-1 and Type-2 AGNs are robustly recovered by the median values and uncertainties of the parameter posteriors estimated by AGNfitter.

4.4. Integrated Luminosities

Integrated luminosities are estimated in order to probe the total power produced by a given radiative process. AGNfitter presents a set of functions, which can be easily adjusted by the user to obtain the preferred integration ranges. In the default version of AGNfitter, we calculate four integrated luminosities over frequency ranges, which are key to the decomposition of host galaxy/AGN contributions (L_{bbb} , L_{gal} , L_{sb} , L_{tor}), as well as two integrated luminosities, which can be converted to key parameters for the physical description of the sources (L_{bol} , L_{IR}). Consistent with our Bayesian approach, AGNfitter computes the total posterior PDFs of the integrated luminosities and returns their median values and percentiles as representative values. Since integration algorithms are computationally expensive, the chains can be additionally “thinned” (e.g., by a factor k), which means we discard all but every k -th observation with the goal of reducing the computational time.

$L_{\text{bbb},0.1-1\mu\text{m}}$, $L_{\text{gal},0.1-1\mu\text{m}}$: we calculate the integrated accretion disk luminosity (BBB) and the integrated host galaxy emission in the frequency range from $0.1-1\mu\text{m}$, in the $\log\nu-\log(\nu L_\nu)$ frame. This interval is chosen due to the simultaneous emission of the AGN and host galaxy, which allows for the study of their relative contribution.

$L_{\text{tor},1-40\mu\text{m}}$, $L_{\text{sb},1-40\mu\text{m}}$: several studies have shown that the mid-infrared regime offers a good laboratory for the study of the ratio of AGN and galaxy contribution to total luminosity (e.g., Lacy et al. 2004; Stern et al. 2005; Donley et al. 2012; Ciesla et al. 2015). AGNfitter computes MIR-luminosities for both the torus and cold dust (starburst) emission components integrating within the relevant wavelength range of $(1-40\mu\text{m})$.

$L_{\text{bol},0.1\mu\text{m}-1\text{KeV}}$: an important probe for the intrinsic power of an AGN is its bolometric luminosity, which we calculate integrating the rest-frame AGN direct emission (BBB) within the frequency interval from $0.1\mu\text{m}$ to 1keV , in the $\log\nu-\log(\nu L_\nu)$ frame. We only consider the BBB template and omit the torus template for the integration, since nearly all photons emitted at infrared wavelengths are reprocessed optical/UV/soft X-ray photons emitted into other directions and re-radiated isotropically. Thus, adding the torus emission to the L_{bbb} calculation would amount to double counting these contributions to the luminosity (e.g., Marconi et al. 2004; Lusso et al. 2012). The X-ray emission would also contribute $\sim 20\%$ to the total bolometric budget, but we omit this contribution since the fitting of X-ray data is not currently implemented in AGNfitter.

$L_{\text{IR},8-1000\mu\text{m}}$: after correcting for any torus contamination, the $8-1000\mu\text{m}$ luminosity is calculated integrating the cold dust emission (starburst component emission). This infrared luminosity is used to estimate the SFR_{FIR} as explained in detail in Section 4.5.

4.5. Host Galaxy Properties

AGNfitter estimates galaxy properties such as stellar masses and SFRs and calculates the total posterior PDFs for these parameters. As before, the median values and uncertainties (16th and 84th percentiles) are given as representative values. The total stellar masses are computed from the masses of the SSP that determine each galaxy template. The normalization of the templates give us a constraint about the number of SSPs and in turn the total stellar mass producing the galactic optical emission.

AGNfitter makes two independent SFR diagnoses both in the optical/UV and the far-infrared (FIR). In the optical regime, the normalization log parameter for the galaxy, G_A , permits to scale the emission of the generic galaxy template to the total emission needed for the fit. With the inferred total stellar mass and taking into account the age and the timescale of the star-formation history of the stellar populations (τ), we can calculate the SFRs according to

$$\left(\frac{\text{SFR}_{\text{opt/UV}}}{M_{\odot} \text{ yr}^{-1}} \right) = 10^{G_A} \times \left(\frac{M_{*sp}}{M_{\odot}} \right) \times \left(\frac{\exp(-age/\tau)}{\tau \text{ yr}^{-1}} \right), \quad (9)$$

where M_{*sp} is the mass of the stellar population.

The total FIR emission provides a different measurement of the SFR since it probes the emission of star-forming regions reprocessed by dust grains in their surroundings. This estimation is of special relevance at redshifts $1 \leq z \leq 3$, where the highest SFR galaxies appear to be dominated by dusty starbursts with high infrared luminosities (Chary & Elbaz 2001; Bouwens et al. 2009). In order to derive the SFR from the starburst model component that dominates the FIR, we use the calibration derived in Murphy et al. (2011) for the total infrared wavelength range of (8–1000 μm),

$$\left(\frac{\text{SFR}_{\text{IR}}}{M_{\odot} \text{ yr}^{-1}} \right) = 3.88 \times 10^{-44} \left(\frac{L_{\text{IR}}}{\text{erg s}^{-1}} \right). \quad (10)$$

Here, L_{IR} is the luminosity integrated over the starburst template in the wavelength range noted above, after a proper subtraction of the AGN torus contribution from the total IR luminosity.

AGNfitter does not include the assumption of energy balance between the direct stellar emission in the optical and the reprocessed emission by cold and warm dust in the IR. This is chosen consciously in order to prevent mixing-up possible data uncertainties of the optical and the FIR. Both components are thus computed independently and this allows us to test the power of each wavelength regime in tracing the SFR. A good agreement between both diagnoses is expected, though the far-infrared SFR tracer is highly dependent on the availability of good FIR data. In the cases where the data set does not contain enough FIR detections, the MIR emission is associated entirely with the torus component. In this way, the starburst emission, and consequently the SFR_{FIR} , are highly underestimated showing no agreement with the optical tracer.

4.6. Spectral Energy Distributions

The probabilistic approach of AGNfitter can be better visualized through the plotting of several different SED realizations, constructed from values randomly picked from the full posterior PDFs of the parameters. In the right panels of Figures 2 and 3, a representative number of SED realizations

are shown simultaneously to exemplify the dynamic range of individual template models and total SEDs contained in the probability distributions. The number of displayed realizations is eight by default but can be freely chosen by the user. Each realization comprises a set of five lines of different colors; one red line that represents the total SED as the linear combination of the physical components, and four lines of different colors, which correspond to the different physical components, following the color coding of Figure 1. Through the variety of SED shapes among single components, it is easy to recognize which components are described by parameters with larger uncertainties. This is clear from the output-SED fitted for the mock Type2 AGN in Figure 3. There the BBB SED appears to be subject to large uncertainties due to a degeneracy among parameters determining the BBB amplitude (BB) and shape (altered by extinction parameter $E(B - V)_{\text{bbb}}$) and the galaxy normalization parameter. Similarly, the parameters determining the starburst template and torus appear to be slightly degenerate, producing a large dynamic range of SED shapes for these components. The degeneracy among these parameters was discussed previously in the context of Figure 5, where the 2D-PDFs of the parameters for the same source were introduced. The visualization of different SED shapes from PDF realizations stresses the relevance of a careful consideration of the parameter uncertainties with the aim of a proper physical interpretation of the SED decomposition.

5. A TEST SAMPLE: XMM-COSMOS

AGNfitter is a multi-purpose SED-fitting tool. It can be applied to different kinds of sources such as quiescent or starburst galaxies, as well as AGNs/quasars, given a source redshift and photometric data. In the rest of this paper, we test the performances of AGNfitter by applying it to both obscured and unobscured X-ray selected AGNs from the XMM-COSMOS survey, taking advantage of the unique multi-wavelength coverage of the COSMOS field (Scoville et al. 2007).

The XMM-COSMOS catalog comprises 1822 point-like X-ray sources detected by XMM-Newton over an area of $\sim 2 \text{ deg}^2$ for a total of $\sim 1.5 \text{ Ms}$ at a fairly homogeneous depth of $\sim 50 \text{ ks}$ (Hasinger et al. 2007; Cappelluti et al. 2009). All the details about the catalog are reported in Brusa et al. (2010). We consider in this analysis 1577 X-ray selected sources for which a reliable optical counterpart can be associated (see discussion in Brusa et al. (2010), Table 1) and we restrict our sample to sources in the catalog with secure spectroscopic redshifts.⁹

5.1. Type1 and Type2 AGN Samples

For the Type-1 AGN sample, we have selected 1375 X-ray sources detected in the [0.5–2] keV band at a flux larger than $5 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ (see Brusa et al. 2010). In this study, we consider 426 objects from this sample that are spectroscopically classified as broad-line AGNs on the basis of broad emission lines ($\text{FWHM} > 2000 \text{ km s}^{-1}$) in their optical spectra (see Lilly et al. 2007; Trump et al. 2009). The origin of spectroscopic redshifts for the 426 sources is as follows: 118 objects from the SDSS archive (Kauffmann et al. 2003;

⁹ The multi-wavelength XMM-COSMOS catalog can be retrieved from http://www.mpe.mpg.de/XMMCosmos/xmm53_release/, version 2011 November 1. This is an updated release of the catalog already published by Brusa et al. (2010).

Adelman et al. 2005), 55 from MMT observations (Prescott et al. 2006), 66 from the IMACS observation campaign (Trump et al. 2007), 130 from the zCOSMOS bright 20k sample, 49 from the zCOSMOS faint catalog (see Lilly et al. 2007), and 8 from individual Keck runs (Brusa et al. 2010). For a detailed description of the sample properties, see Elvis et al. (2012) and Hao et al. (2012).

To select an obscured AGN sample, we have instead considered [2–10] keV detected objects in the *XMM*-COSMOS catalog having fluxes larger than $3 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ (971 sources). From this sample, we selected 288 AGNs with spectroscopic redshift and optical spectra lacking broad emission lines (i.e., $\text{FWHM} < 2000 \text{ km s}^{-1}$): 259 are objects with either spectrally unresolved, high-ionization emission lines, exhibiting line ratios indicating AGN activity, or not detected high-ionization lines, where the observed spectral range does not allow us to construct line diagnostics (Brusa et al. 2010). Not all AGNs selected via this criterion show emission lines, e.g., 29 are classified absorption-line galaxies, i.e., sources consistent with a typical galaxy spectrum showing only absorption lines. Further details on the obscured AGN sample and on their properties are given by Lusso et al. (2012)

Altogether, the sample used for testing the capabilities of AGNfitter consists of 714 X-ray selected active galaxies optically classified into 426 Type1 and 288 Type2 AGNs. Due to Malmquist bias and owing to the different selection criteria, the redshift distributions of the two populations are considerably different, so that Type1 AGNs cover a redshift range between $0.10 < z < 4.26$ ($\langle z_{T1} \rangle = 1.64$) while Type2s present redshifts in the range of $0.05 < z < 3.52$ ($\langle z_{T2} \rangle = 0.85$).

5.2. Multi-wavelength Coverage

The catalog includes multi-wavelength data from far-infrared to hard X-rays: *Herschel* data at 160 and $100 \mu\text{m}$ (Lutz et al. 2011), 70 and $24 \mu\text{m}$ MIPS GO3 data (Le Flocc’h et al. 2009), IRAC flux densities (Ilbert et al. 2010), near-infrared J UKIRT (Capak et al. 2008), H-band (McCracken et al. 2010), CFHT/K-band data (McCracken et al. 2008), optical multiband photometry (SDSS, Subaru, Capak et al. 2007), and near- and far-ultraviolet bands with *GALEX* (Zamojski et al. 2007). The observations in the optical–UV and near-infrared bands are not simultaneous because they span a time interval of about five years: 2001 (SDSS), 2004 (Subaru and CFHT), and 2006 (IRAC). In order to reduce possible variability effects, we have selected the bands closest in time to the IRAC observations (i.e., we excluded SDSS data, which in any case are less deep than other data available in similar bands).

Galactic reddening has been taken into account: we used the selective attenuation of the stellar continuum $k(\lambda)$ taken from Table 11 of Capak et al. (2007). Galactic extinction is estimated from Schlegel et al. (1998) for each object.

5.3. Treatment of Upper Limits

The COSMOS photometric catalogs do not provide forced photometry at the location of a source in the bands for which the object flux lies below the formal detection limit. Instead, only upper limits are reported. These upper limits can nevertheless be highly informative and place important constraints on model parameters, but we must consider how to implement these upper limits into our Gaussian likelihood. The correct approach would obviously be to perform forced

photometry in all bands where a formal detection is not provided (Lang et al. 2014), which would give us a measurement and an error, with a large error reflecting the fact that the measurement (a marginal or non-detection) is very noisy. However, adding forced photometry to the entire *XMM*-COSMOS AGN sample is beyond the scope of the present work, hence we adopt the following crude approximation, which nevertheless allows us to incorporate the information provided by the COSMOS upper limits. For non-detections, where the COSMOS catalog reports 5σ flux upper limits (F_+), AGNfitter creates a fictitious data point at a flux level of $0.5F_+$, and a fictitious symmetric error bar of $\pm(0.5F_+)$. In this way, upper limits can be trivially incorporated into our Gaussian likelihood in Equation (4). Because our “upper limit” region spans from zero to F_+ , this approach allows our MCMC sampling to accept all models with fluxes lying between zero and the upper limit ones. The inclusion of upper/lower limits in this fashion is necessary, since the information they provide can be very helpful for constraining the likelihood of some models.

6. RESULTS AND DISCUSSION ON THE TEST SAMPLE

6.1. Spectral Energy Distributions for Type1 and Type2 AGNs

In this section, we discuss general properties within the Type1 and Type2 AGN populations from an SED perspective. In Figure 6, we present some examples of SED fitting for Type1 AGNs having low levels of extinction in the optical–UV ($E(B - V)_{\text{bb}} < 0.2$). To visualize the dynamic range of the parameter values included in the PDF, we randomly pick eight different realizations from the posterior PDFs of each corresponding component SED. Since the AGN sample we considered in our analysis is selected in the X-rays, it is less prone to be biased in the optical. This gives us the opportunity to test AGNfitter over a wider variety of AGN/galaxy contributions than optically selected ones.

XID = 21 and 5257 are good examples of how an AGN SED looks in the case of both low-reddening and low-host galaxy contamination. Overall, 58% (248/426) of the total Type1 AGN sample has $E(B - V)_{\text{bb}} \lesssim 0.2$. The BBB and the torus properties for these objects are nicely constrained and a clear dip at $1 \mu\text{m}$ is visible, marking the division between hot-dust and accretion disk emission, consistent with what it is typically found in optically selected AGN samples (Elvis et al. 1994; Richards et al. 2006a; Shang et al. 2011; Stern & Laor 2012). On the other hand, for these sources the fits to the the host galaxy and cold-dust components are not very informative. For the former, this is because the AGN significantly outshines its host, whereas the latter is simply due to a lack of FIR detections.

There are also cases for which, although the BBB emission is significant ($L_{\text{bbb,dered}} > 10^{45} \text{ erg s}^{-1}$), the galaxy contribution at $\sim 1 \mu\text{m}$ is not completely negligible, making the overall shape of the AGN rather flat over several decades in frequency (see XID = 21, 52, and 5081 as examples). As the BBB luminosity of the source decreases ($L_{\text{bbb,dered}} \simeq 10^{44-45} \text{ erg s}^{-1}$), even for cases with low BBB reddening, some sources require a significant galaxy contribution at $\sim 1 \mu\text{m}$ (e.g., XID = 13, 53283, and 54513). About 75% of the sample has a contribution of the host galaxy of about 50% or more at $1 \mu\text{m}$. The lack of a $1 \mu\text{m}$ inflection point in the SED between the UV and near-IR bumps (i.e., a conspicuous host galaxy contribution) is a

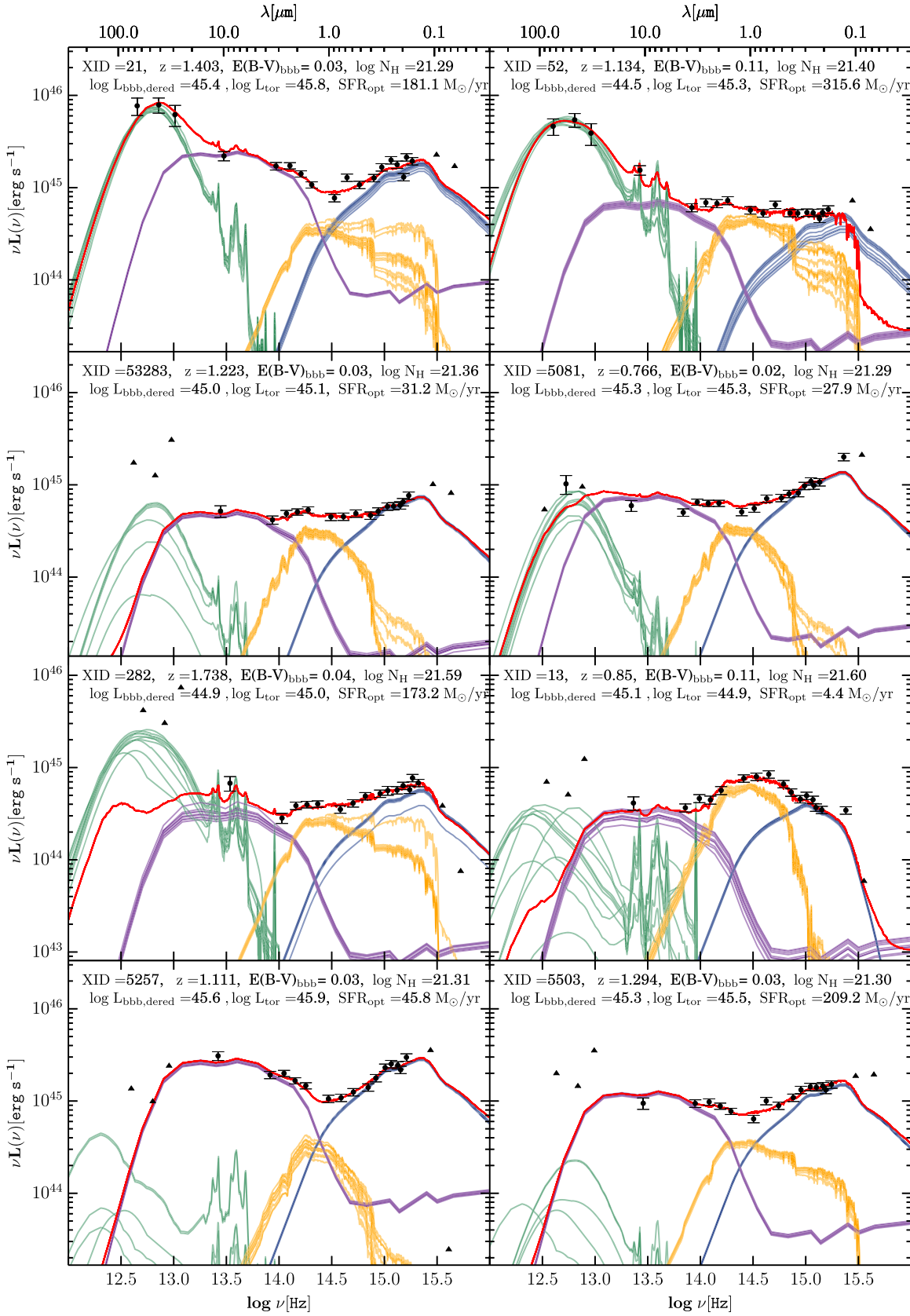


Figure 6. Example SEDs for Type 1 AGNs: photometric detections are plotted against rest-frame wavelengths/frequencies as circular markers with error bars, while non-detections are represented as triangles. SED shapes of the physical components are presented as solid lines with colors assigned similarly to Figure 1, while the linear combination of these, the “total SED,” is depicted as a red line. We pick eight different realizations from the parameter posterior probability distributions and over-plot these SEDs in order to visualize the effect of the parameters’ uncertainties on the SEDs.

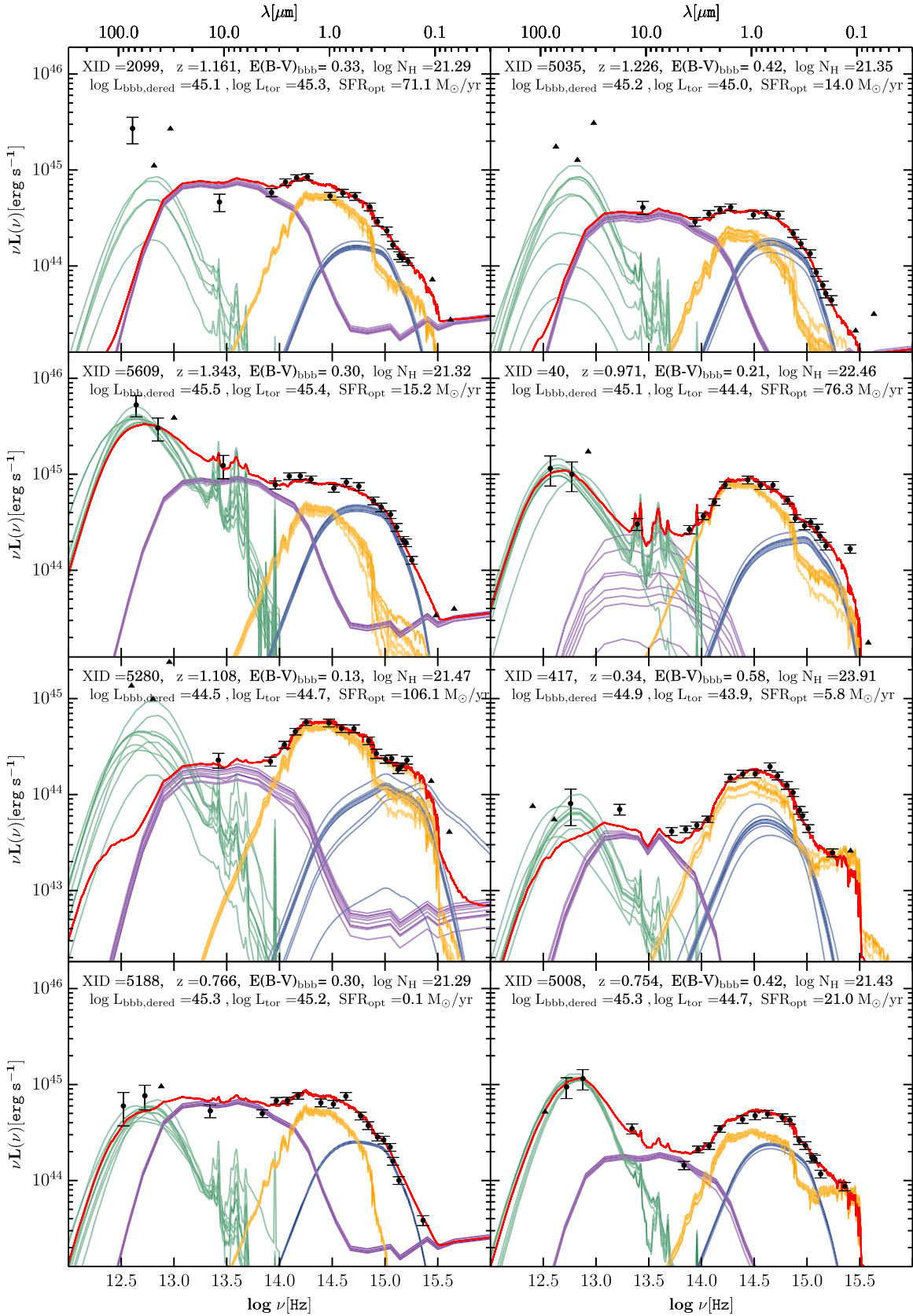


Figure 7. Example SEDs for reddened Type 1 AGNs: Photometric detections are plotted against rest-frame wavelengths/frequencies as circular markers with error bars, while non-detections are represented as triangles. SED shapes of the physical components are presented as solid lines with colors assigned similarly to Figure 1, while the linear combination of these, the “total SED,” is depicted as a red line. We pick eight different realizations from the parameter posterior probability distributions and over-plot these SEDs in order to visualize the effect of the parameters’ uncertainties on the SEDs.

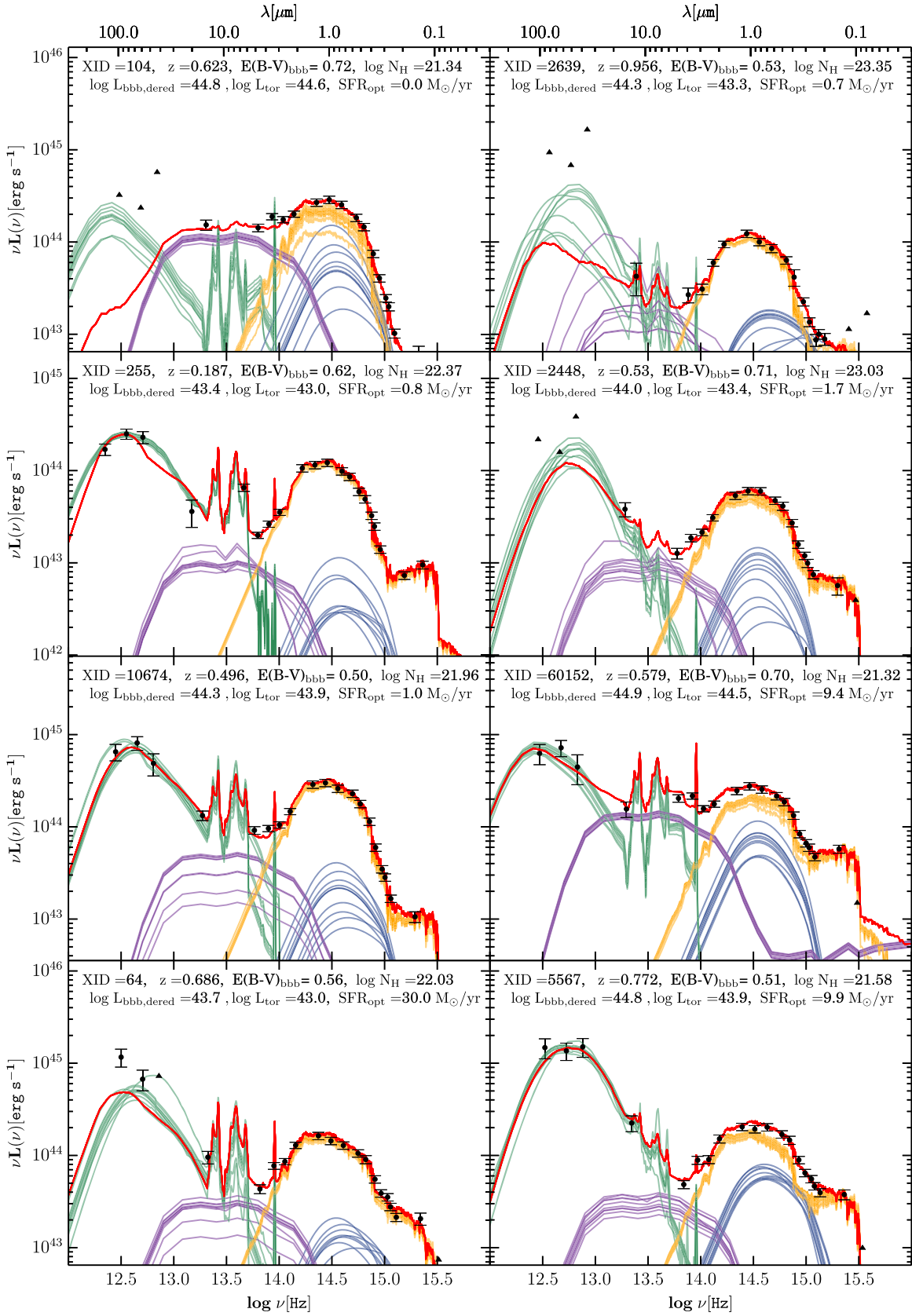


Figure 8. Example SEDs for Type 2 AGNs: photometric detections are plotted against rest-frame wavelengths/frequencies as circular markers with error bars, while non-detections are represented as triangles. SED shapes of the physical components are presented as solid lines with colors assigned similarly to Figure 1, while the linear combination of these, the “total SED,” is depicted as a red line. We pick eight different realizations from the parameter posterior probability distributions and over-plot these SEDs in order to visualize the effect of the parameters’ uncertainties on the SEDs.

common feature of X-ray selected AGN samples, since these usually contain fainter AGNs for which the contrast to galaxy emission is lower (Elvis et al. 2012; Hao et al. 2014).

Figure 7 presents some examples of Type1s with significant amount of reddening in the optical–UV ($E(B - V)_{\text{bb}} \gtrsim 0.2$; about 42% of the Type1 AGN sample). The SEDs of such sources (e.g., XID = 2099, 5035, 5609, and 40) are characterized by a sharp decline of the BBB in the optical–UV ($\lambda \sim 0.5 - 0.2 \mu\text{m}$), and they can be considered AGNs of intermediate-type ($\sim 1.5 - 1.9$). Objects with appreciable reddening still show degeneracy between BBB and the host galaxy, despite our implementation of a prior on the maximum galaxy luminosity allowed at a given redshift. The BBB of a moderately reddened Type 1 AGN can also be equivalently modeled with a bright galaxy having very young stellar ages and a BBB with high reddening values (e.g., see XID = 417, 5280 in Figure 7 as examples), and as expected this degeneracy is even more pronounced in the case of low-luminosity AGNs ($L_{\text{bol}} \sim 10^{44} \text{ erg s}^{-1}$). To break this degeneracy, an independent measurement of the host galaxy emission is needed, though it is not trivial to estimate for unobscured AGNs. One possibility is either to increase the threshold defining the characteristic galaxy luminosity (as discussed in Section 3.5), and/or to include spectral information. For example, high $H\alpha/H\beta$ ratio ($\gg 3$) indicates high reddening (Dong et al. 2008).

Regarding mid-IR properties, Type1s are preferentially fitted with torus templates corresponding to an obscuring media of considerably low hydrogen column densities ($\log N_{\text{H}} < 22$). A quantitative study of the obscuration properties of this subsample is given in Section 6.3.

Figure 8 shows eight examples of SEDs for spectroscopically classified Type2s. Type2 AGN SEDs are dominated by the host galaxy emission at $0.5 - 2 \mu\text{m}$, while the BBB component is obviously more difficult to constrain, spanning a wider dynamic range than in the Type1 cases. Type2s are also characterized by high BBB reddening values (91% have $E(B - V)_{\text{bb}} \gtrsim 0.2$), as well as large column densities of the torus component. Obscuration properties will be discussed in detail in Section 6.3. Overall, the optical–UV portion of the Type-2 SEDs is very well fitted by stellar emission, with negligible contribution of the BBB. We also point out that the near-infrared emission at $10 - 20 \mu\text{m}$ might be contaminated by emission features due to the stochastic heating of PAH molecules or carbon grains (e.g., Flagey et al. 2006). These features are associated with massive star-forming regions at much lower dust temperature ($T < 100 \text{ K}$). PAHs can be non-negligible in obscured AGNs and give rise to possible degeneracies with the torus emission produced by dust at parsec scales. Such degeneracies may be solved by including priors on the cold-dust/torus emission, which take into account the near-IR spectral information if available (e.g., line ratios, silicate absorption feature at $9.7 \mu\text{m}$).

It is reasonable to ask whether the physical properties inferred by AGNfitter allow us to robustly distinguish between unobscured and obscured AGNs. In the following discussion, we will use the obscuration properties inferred by AGNfitter to re-classify the total sample into Type1s and Type2s and we will compare our classification to the spectroscopic one available in XMM-COSMOS.

6.2. Physical Parameters

Figure 9 presents the output of AGNfitter on physical parameters for the total XMM-COSMOS AGN sample. As shown here, AGNfitter delivers vast information about integrated AGN luminosities, properties of the host galaxies, and obscuration parameters. The large number of inferred properties can motivate a variety of science-cases related to AGNs.

Figure 9 shows the parameter PDFs for the total Type1 (blue histograms) and Type2 AGN populations (orange histograms). These histograms were constructed sampling 500 random draws from each source’s posterior PDF. Distinct from previous work, our treatment thus enables a proper inclusion of the parameter uncertainties in the histograms, allowing a fully probabilistic study of the global properties of the two populations under consideration, and other inference effects, which may shape the observed parameter PDFs.

The upper four histograms of Figure 9 show the integrated luminosity distributions of Type1s and Type2s for $\log L_{\text{tor}}$, $\log L_{\text{bbb-dered}}$, $\log L_{\text{gal}}$, and $\log L_{\text{bbb-red}}$. On average, Type1s appear to be more luminous than Type2s in all four parameters, but this is clearly a selection effect because the average redshifts probed by these two samples are different, i.e., Type1s are, on average, observed at higher redshifts than Type2s (see Section 5). For instance, the $\log L_{\text{tor}}$ median values for Type1 AGNs are, on average, over an order of magnitude higher ($45.20^{+0.61}_{-0.75}$) than those for Type2s ($44.01^{+0.98}_{-0.92}$).

The distribution of $L_{\text{bbb-red}}$ for Type1 AGNs has a median value of $\log L_{\text{bbb-red}} = 44.55^{+0.83}_{-0.83}$ (the quoted uncertainties represent the 16th and the 84th percentiles of the distribution), while the median value for Type2s is $\log L_{\text{bbb-red}} = 43.59^{+0.62}_{-0.57}$, with a long tail to low BBB luminosities. Once we correct for extinction, the two $\log L_{\text{bbb-dered}}$ distributions nearly overlap within the uncertainties centered on similar median luminosity ranges: $45.08^{+0.65}_{-0.68}$ and $44.56^{+0.86}_{-0.61}$ for Type1s and Type2s, respectively, consistent with previous analyses of X-ray selected AGN samples (Lusso et al. 2011, 2012, 2013; Bongiorno et al. 2012). 58% of the Type1 AGNs appear to have relatively low levels of extinction ($E(B - V)_{\text{bb}} < 0.2$), while the majority of Type2s exhibit significant reddening (91% with $E(B - V)_{\text{bb}} \geq 0.2$), meaning a more appreciable reddening correction for the latter with respect to the Type1s. As a result, the shift of the $\log L_{\text{bbb-dered}}$ distribution for Type2s to higher values is more pronounced than for Type1s. AGN obscuration properties related to the hydrogen column density parameter N_{H} will be further discussed in Section 6.3.

On the other hand, the starburst luminosities ($\log L_{\text{sb}, 1-40 \mu\text{m}}$) show a rather flat and uninformative distribution for Type1 AGNs with respect to Type2s. A similar result is obtained for the FIR SFRs (SFR_{IR} ; which are derived from the total infrared emission $L_{\text{IR}, 8-1000 \mu\text{m}}$). This is mainly due to the higher fraction of *Herschel* non-detections in the Type1 AGN with respect to the obscured AGN population. Only $\sim 10\%$ of Type1 AGNs are detected at both 100 and 160 μm , while Type2s have a detection fraction about three times higher (whereas at 70 μm , $\sim 8\%$ and 15% of Type1 and Type2 are detected respectively). We find that the median L_{sb} luminosities and SFRs for Type2 AGNs are $\log L_{\text{sb}} = 43.80^{+0.85}_{-1.14}$ and $\log \text{SFR}_{\text{IR}} = 0.95^{+0.82}_{-1.12}$, in agreement with previous analyses

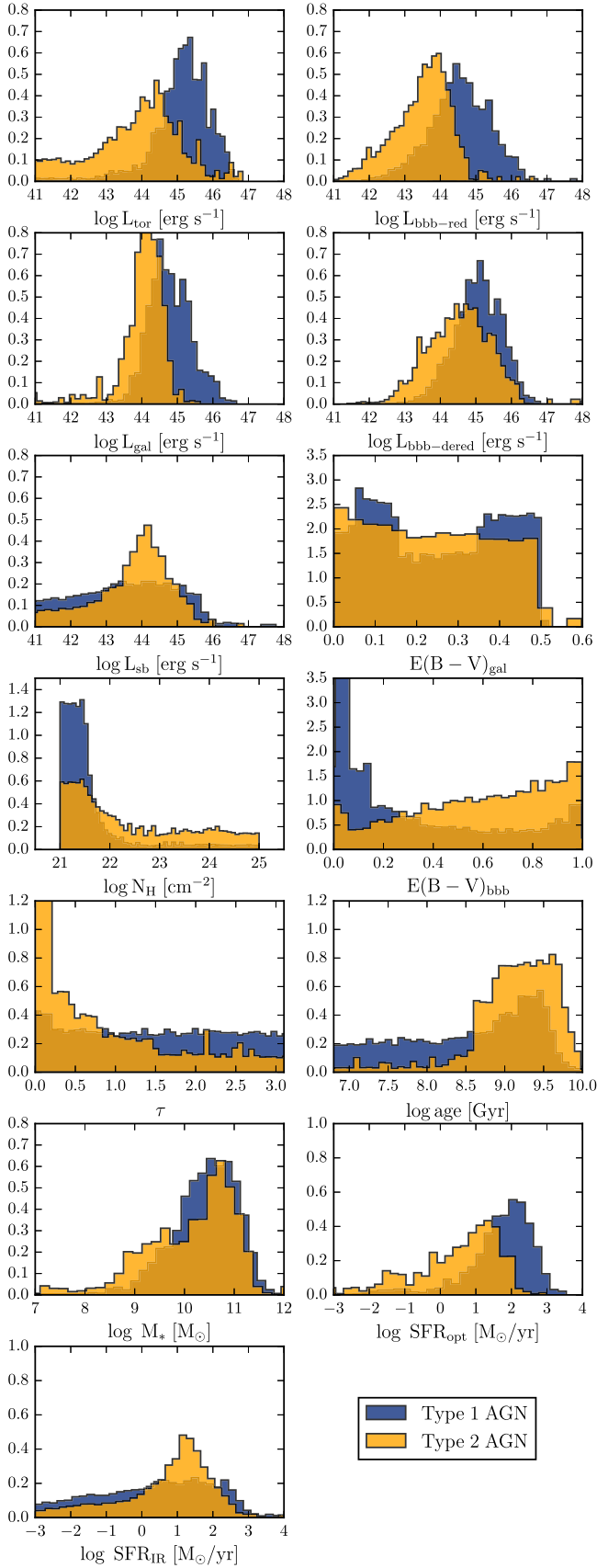


Figure 9. Distribution of the output parameters for the whole Type1 (blue histogram) and Type 2 (orange histogram) AGN population. These histograms were constructed, sampling 500 random draws from each source’s PDFs of each parameter.

(Lusso et al. 2011; Bongiorno et al. 2012). We conclude that at least one detection in the FIR bands is essential for any analysis of the cold dust properties of the galaxy. Such a study will be presented for a sample of radio-selected AGNs and star-forming galaxies in a forthcoming paper (G. Calistro Rivera et al. 2016, in preparation).

Previous studies found X-ray selected AGNs to reside preferentially in massive ($M_* > 10^{10} M_\odot$) bulge-dominated galaxies with red colors (e.g., Lusso et al. 2011; Bongiorno et al. 2012, but see also Silverman et al. 2009; Xue et al. 2010). On average, stellar masses for Type1 AGNs appear to be slightly larger than that of Type2s; though, their difference is not statistically significant within their uncertainties ($\log M_* = 10.46^{+0.56}_{-0.73}$ and $\log M_* = 10.29^{+0.63}_{-0.82}$ for Type1s and Type2s, respectively). Due to their galaxy dominated SED shapes, Type2s have really well constrained host galaxy physical parameters, showing relatively older stellar populations (age = $1.51^{+2.69}_{-0.80}$ Gyr) with respect to the Type1s (age = $0.27^{+2.02}_{-0.264}$ Gyr), in agreement with earlier works (Lusso et al. 2011; Bongiorno et al. 2012). Type1s have instead a much wider distribution of both τ and ages, which means that recovering the host galaxy properties for unobscured AGNs is a challenging task, subject to considerably larger uncertainties than for Type2s. In general, stellar masses are robustly constrained despite of the presence of a luminous AGN, since these are primarily determined by the emission at around rest frame $1 \mu\text{m}$, where the contrast between the BBB and galaxy is usually minimized (Stern & Laor 2012). This is however not the case for the SFR parameter, since it is highly sensitive to the optical/UV SED, which is on its turn highly sensitive to AGN contribution, especially for Type1 AGNs.

For completeness, we also reported in Figure 9 the distributions for the host galaxy reddening $E(B - V)_{\text{gal}}$, and optical SFRs (SFR_{opt} , corrected for extinction) for both AGN populations; though, we note that the reliability of UV-based SFR indicators in galaxies has long been debated (e.g., Adelberger & Steidel 2000; Bell 2002).

All in all, due to the different redshift ranges covered by the two AGN populations, ($\langle z_{\text{T2}} \rangle \sim 0.85$, $\langle z_{\text{T1}} \rangle \sim 1.64$), one should consider the effect of the redshift while interpreting the results to avoid comparing sources that may be at different evolutionary stages. Nonetheless, our main aim here is to provide an overall view of the full range of the physical parameters estimated by AGNfitter for the AGN population as a whole.

6.3. AGN Obscuration and Classification

The parameters directly related to the obscuration properties of the nuclear emission are the dust reddening parameter $E(B - V)_{\text{bbb}}$ and the column density N_{H} . The parameter $E(B - V)_{\text{bbb}}$ quantifies the absorption and reprocessing of the direct AGN emission by gas and dust along the line of sight at host galaxy scales, while the column density parameter ($\log N_{\text{H}}$) describes the absorbing dust distributed mainly at nuclear scales.

We compare AGNfitter results on these parameters with those yielded by the non-Bayesian SED-fitting code presented in Lusso et al. (2013) for the Type1 population. While they find a median dust reddening value of $E(B - V)_{\text{bbb}} \leq 0.03$, AGNfitter finds a median value of $E(B - V)_{\text{bbb}} = 0.13^{+0.58}_{-0.11}$ (see Figure 9). The deviation from the literature results can be explained through the difference between our estimated median

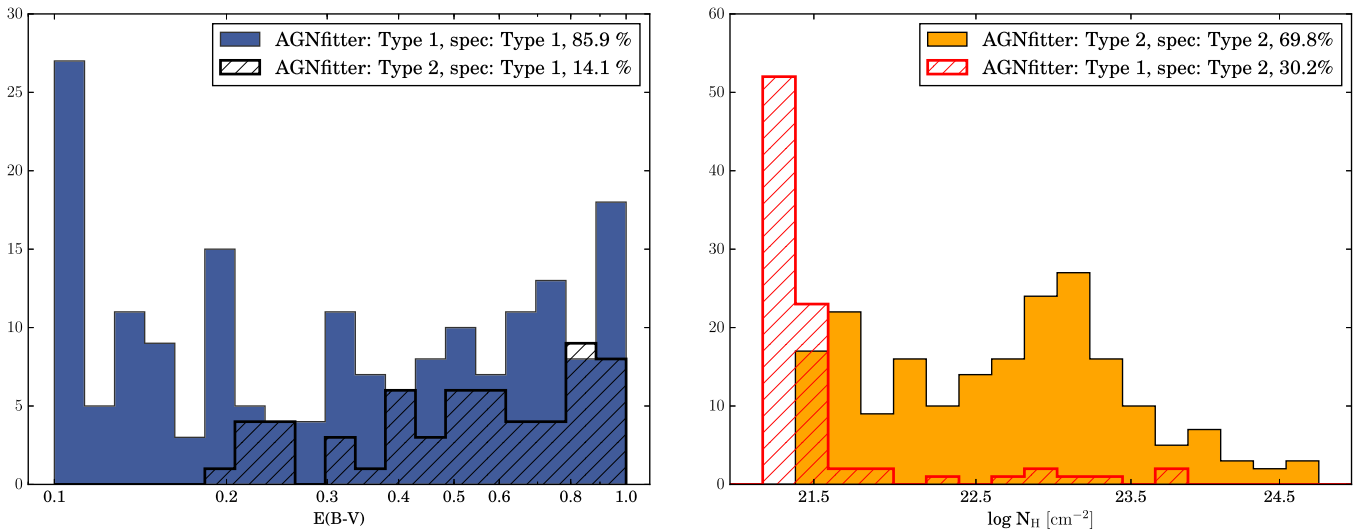


Figure 10. AGNfitter vs. spectroscopic classification: the left panel shows the distribution of the Type1 sources for the reddening parameter $E(B - V)_{\text{bbb}}$. The filled blue distribution represents the spectroscopically classified Type1s that are also classified as such in AGNfitter, while the dashed black distribution represents the ones misclassified as Type2s. The right panel shows the distribution of the Type2 sources for the column density parameters $\log N_{\text{H}}$. The filled orange distribution represents the spectroscopically classified Type2s that are also classified as such in AGNfitter, while the dashed red distribution represents spectroscopic Type2s misclassified as Type1s by AGNfitter.

values from the PDFs and the maximum likelihood ones (best-fit) usually adopted in the literature.

This difference is especially relevant in the cases where the parameter $E(B - V)_{\text{bbb}}$ has an almost flat posterior PDF (as is the case for 36% of the Type1 AGNs, which have a total error in $E(B - V)_{\text{bbb}}$ larger than 0.2). For this kind of Type 1 AGN, the maximum likelihood estimate of the reddening parameter is not precise and the flat and broad PDF pushes the median value to larger reddening values, which also results in large uncertainties on this parameter (see Assef et al. 2010 for similar conclusions).

In contrast to the Type1s, no more than 9% of the Type2 subsample present low-reddening values of $E(B - V)_{\text{bbb}} \leq 0.1$. The large majority of Type2 are highly reddened sources with a population median value of $E(B - V)_{\text{bbb}} = 0.64^{+0.28}_{-0.25}$.

A similar behavior can be observed in the results for the column density parameter, $\log N_{\text{H}}$. As can be seen in Figure 9, Type1 AGNs clearly prefer low column densities with $\sim 86\%$ having $\log N_{\text{H}} < 22$ as expected from their unobscured emission lines (population median value $N_{\text{H}} = 21.39^{+0.51}_{-0.26}$). Type2 AGNs on the contrary, are more spread along large regimes of low and high column density, with only $\sim 45\%$ having $\log N_{\text{H}} < 22$ (population median value $N_{\text{H}} = 22.02^{+1.95}_{-0.58}$). These column density values result from the different shapes of the torus SEDs, where higher column density is characterized by less near-IR emission of shorter wavelength, probably due to self-absorption of this energetic dust emission by the dust at lower temperatures distributed at larger scales (e.g., Figure 1). Overall, AGNfitter is able to recover general obscuration properties of both AGN classes.

The slight bimodality in the distributions for these two independent obscuration parameters provides the opportunity to develop a classification strategy. A multi-wavelength classification through these AGNfitter parameters provides the clear advantage of being sensitive to photometric measurements over a wide wavelength range (mid-IR to opt-UV), in contrast to a spectral classification, which covers only a few

thousand Angstroms in the rest frame. Moreover, while spectroscopy can be rather expensive and is not always available for all sources, our method can be easily applied to complete multi-wavelength datasets.

We define the classification of the two AGN populations to be the following.

Type1 AGN:

$$N_{\text{H}} < 21.5 \cup 0 < E(B - V)_{\text{bbb}} < 0.2.$$

Type2 AGN:

$$N_{\text{H}} > 21.5 \cap E(B - V)_{\text{bbb}} > 0.2.$$

Although this is a hybrid sample because of the different selection effects employed, it nevertheless provides useful estimates for the true completeness and efficiency of classification for more uniformly selected X-ray AGN samples. The results of our classification scheme are presented in Figure 10. As can be seen in the left panel, the majority of the broad emission line AGNs spectroscopically classified as Type1s are also classified as Type1s by the AGNfitter obscuration parameters, proving a completeness of (366/426, $\sim 86\%$). For the case of Type2s in the right panel of Figure 10, our classification method is complete in $\sim 70\%$ (201/288). A small fraction of the spectroscopically classified Type1s are misclassified as Type2s by AGNfitter's method (89/455), showing to be efficient at $\sim 80\%$ for Type1 AGNs. The false-positive ratio is slightly larger for Type2s, being 60/261 of spectroscopically classified Type2s misclassified as Type1 AGNs by our method, proving an efficiency of $\sim 77\%$ for Type2 sources.

The cases of disagreement with the spectroscopic approach does not necessarily imply that the SED-fitting classification is incorrect. On the contrary, multi-wavelength approaches as AGNfitter can be more sensitive to the global properties of the sources for such a classification, since it analyzes the obscuration level in different independent components of the SED.

A similar approach has been used by Assef et al. (2013) based on a non-Bayesian SED-fitting code, where they adopt $E(B - V) = 0.15$ as the dividing line for their classification.

They choose this classification using an estimate from the standard X-ray boundary of a gas column density of $N_H = 10^{22} \text{ cm}^{-2}$ (Ueda et al. 2003) and the median value of the ratio $E(B - V)/N_H = 1.5 \times 10^{-23} \text{ cm}^2 \text{ mag}$ observed for the sample presented in Maiolino et al. (2001). Although a direct completeness comparison to the method of Assef et al. (2013) is not possible, since they do not test their method on spectroscopic classification, we calculate the completeness of their classification strategy applied to the fitting results of our sample. Following Assef et al. (2013) and using $E(B - V) = 0.15$ as the dividing line for the classification, we obtain a completeness ratio of 54% and 91% and an efficiency of 90% and 57% for Type1 AGNs and Type2 AGNs respectively. In general, our method presents slightly better completeness and efficiency ratios, recovering a significantly larger fraction of Type1 sources, while slightly compromising the completeness of Type2s. As this disagreement is likely to arise from the different SED-fitting approaches used to infer the obscuration parameters, a direct comparison of their classification strategy to spectroscopic classified samples would be needed.

7. SUMMARY AND CONCLUSIONS

We introduced AGNfitter: a fully Bayesian statistical tool for fitting and decomposing SEDs of active galaxies through MCMC sampling of the model parameters. The total active galaxy model in AGNfitter consists of the host galaxy emission, modeled as a combination of a stellar component and a starburst cold gas component and the nuclear AGN emission, modeled as a combination of an accretion disk (BBB) and a hot dust torus component. For both AGN and host galaxy models in the optical and UV, the effect of reddening along the line of sight is accounted for and corrected. Through informed sampling, the algorithm explores the parameter space that defines this model and computes the full PDFs of the parameters. AGNfitter calculates the median values with respective uncertainties of the following parameters:

1. physical parameters: τ , age, $\log N_H$,
2. reddening parameters: $E(B - V)_{\text{bbb}}$ and $E(B - V)_{\text{gal}}$, and
3. normalization variables: SB, BB, GA, and TO.

Multiple relevant properties for the characterization of the AGN and the host galaxy are also calculated. These are, for example, the integrated luminosities of the AGN components (e.g., L_{bol} , L_{bbb} , $L_{\text{bbb-dered}}$ and L_{tor}) and host galaxy properties (e.g., L_{gal} , L_{sb} , M_* , SFR_{opt} , and SFR_{FIR}).

The capability of AGNfitter in properly inferring these source properties was tested on mock active galaxies created as prototypical Type1 and Type2 SEDs. AGNfitter could accurately recover the input parameters of the synthetic SEDs and also provide insights into the degeneracies between emission components through the shapes of their PDFs.

The performance of the code for real data was tested on a sample of X-ray selected AGNs from the *XMM-COSMOS* survey, which provides 15 band photometry from the UV to the far-IR for the construction of the SEDs. AGNfitter was applied to 714 sources (426 Type1 and 288 Type2 AGNs), which were previously spectroscopically classified as Type1 and Type2 sources by their optical emission lines. The fitting results include two independent model parameters, which are proxies for AGN obscuration. This allowed us to develop a

classification strategy for unobscured (Type1) versus obscured (Type2) AGNs, which shows a great agreement with the spectroscopic classification. The completeness fraction of our classification scheme is $\sim 86\%$ and $\sim 70\%$, with an efficiency of $\sim 80\%$ and $\sim 77\%$, for Type1 and Type2 AGNs respectively.

The complexity of AGN physics leaves many possibilities of improving the physical models used to approximate AGN emission. While they might be more detailed and accurate models for the different AGN components, these imply the treatment of larger parameter spaces with an unavoidable increase of degeneracies among the parameters. In those cases, the use of a Bayesian methodology as the one of AGNfitter is even more crucial.

The efficiency and code structure of AGNfitter allows new models to be easily integrated to the existing libraries. One significant improvement would be to implement a more flexible BBB model, which can be based on a physically motivated accretion disk model, possibly dependent on black hole mass and disk accretion rate (e.g., Slone & Netzer 2012), instead of a single empirical template. A further possible change could be the implementation of a more complex prior on the galaxy models to reflect the information from their total luminosity function. An interesting addition would be to allow the user to explore the effect on the galaxy/AGN parameters if an energy balance between the cold-dust and optical-UV star formation is considered (e.g., da Cunha et al. 2008). Moreover, the implementation of a model of the IGM absorption such that rest-frame UV wavelengths at $\lambda < 1250 \text{ \AA}$ can be included in the fitting procedure, would be an additional improvement of our model.

AGNfitter’s multi-wavelength approach allows the simultaneous study of multiple physical processes. This versatility makes of AGNfitter a flexible tool to address several physical questions related to AGNs and galaxies. The AGNfitter python code is available under an open-source MIT license at <https://github.com/GabrielaCR/AGNfitter> and the version described in this paper ([v0.9]) is archived on Zenodo (Calistro Rivera 2016, DOI:10.5281/zenodo.154655).

The authors thank the anonymous referee for a careful reading and useful comments, which have improved this manuscript. Joseph F. Hennawi acknowledges generous support from the Alexander von Humboldt foundation in the context of the Sofja Kovalevskaja Award. The Humboldt foundation is funded by the German Federal Ministry for Education and Research. Gabriela Calistro Rivera gratefully acknowledges support from the European Research Council under the European Unions Seventh Framework Programme (FP/2007-2013)/ERC Advanced Grant NEWCLUSTERS-321271. We thank the members of the ENIGMA group¹⁰ at the Max Planck Institute for Astronomy (MPIA) for insightful suggestions and discussions. We thank Daniel Foreman Mackey for his support on the implementation of the code Emcee as the MCMC core of AGNfitter. We thank Jonathan Stern for useful comments on the manuscript. We thank Neil H. M. Crighton, Jose Onorbe, Alberto Rorai, Kenneth Duncan, Wendy Williams, and Marco Velliscig for important contributions in the scripting of the algorithm in Python. We thank Micol Bolzonella and Elena Zucca for useful comments and suggestions on the implementation of the prior on the galaxy

¹⁰ <http://www.mpia-hd.mpg.de/ENIGMA/>

luminosity function. We thank Huub Röttgering for his support during the final phase of the project.

REFERENCES

- Acquaviva, V., Gawiser, E., & Guaita, L. 2011, *ApJ*, 737, 47
- Adelberger, K. L., & Steidel, C. C. 2000, *ApJ*, 544, 218
- Adelman, S. J., Caliskan, H., Gulliver, A. F., & Teker, A. 2005, *yCat*, 344, 0
- Antonucci, R. 1993, *ARA&A*, 31, 473
- Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, *ApJ*, 713, 970
- Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, *ApJ*, 772, 26
- Bell, E. F. 2002, *ApJ*, 577, 150
- Berta, S., Lutz, D., Santini, P., et al. 2013, *A&A*, 551, A100
- Bongiorno, A., Merloni, A., Brusa, M., et al. 2012, *MNRAS*, 427, 3103
- Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, *ApJ*, 705, 936
- Brauer, J. R. 2003, in *Exploiting the ISO Data Archive. Infrared Astronomy in the Internet Age*, ed. C. Gry et al. (ESA SP-511; Noordwijk: ESA), 21
- Brusa, M., Civano, F., Comastri, A., et al. 2010, *ApJ*, 716, 348
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Calistro Rivera, G. 2016, *GabrielaCR/AGNfitter: Last Beta Release*, v0.9, Zenodo, doi:10.5281/zenodo.154655
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Capak, P., Aussel, H., Ajiki, M., et al. 2007, *ApJS*, 172, 99
- Capak, P., Aussel, H., Ajiki, M., et al. 2008, *yCat*, 2284, 0
- Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, *A&A*, 497, 635
- Chabrier, G. 2003, *PASP*, 115, 763
- Chary, R., & Elbaz, D. 2001, *ApJ*, 556, 562
- Ciesla, L., Charmandaris, V., Georgakakis, A., et al. 2015, *A&A*, 576, A10
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, 388, 1595
- Dale, D. A., & Helou, G. 2002, *ApJ*, 576, 159
- Dale, D. A., Helou, G., Contursi, A., Silbermann, N. A., & Kolhatkar, S. 2001, *ApJ*, 549, 215
- Dong, X., Wang, T., Wang, J., et al. 2008, *MNRAS*, 383, 581
- Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *ApJ*, 748, 142
- Elvis, M., Hao, H., Civano, F., et al. 2012, *ApJ*, 759, 6
- Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, *ApJS*, 95, 1
- Flagey, N., Boulanger, F., Verstraete, L., et al. 2006, in *ASP Conf. Ser. 357. The Spitzer Space Telescope: New Views of the Cosmos*, ed. L. Armus & W. T. Reach (San Francisco, CA: ASP), 85
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Gelman, A., & Rubin, D. B. 1992, *StaSc*, 7, 457
- Goodman, J., & Weare, J. 2010, *Comm. Appl. Math. Comput. Sci.*, 5, 65
- Granato, G. L., & Danese, L. 1994, *MNRAS*, 268, 235
- Han, Y., & Han, Z. 2012, *ApJ*, 749, 123
- Hao, H., Elvis, M., Civano, F., et al. 2012, arXiv:1210.3033
- Hao, H., Elvis, M., Civano, F., et al. 2014, *MNRAS*, 438, 1288
- Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, *ApJS*, 172, 29
- Hastings, W. K. 1970, *Biometrika*, 57, 97
- Hatziminaoglou, E., Fritz, J., Franceschini, A., et al. 2008, *MNRAS*, 386, 1252
- Hernán-Caballero, A., Alonso-Herrero, A., Hatziminaoglou, E., et al. 2015, *ApJ*, 803, 109
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, *ApJS*, 163, 1
- Hopkins, P. F., Strauss, M. A., Hall, P. B., et al. 2004, *AJ*, 128, 1112
- Ilbert, O., Salvato, M., Le Flocc'h, E., et al. 2010, *ApJ*, 709, 644
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, *MNRAS*, 442, 1805
- Iovino, A., Cucciati, O., Scoddeggio, M., et al. 2010, *A&A*, 509, A40
- Johnson, S. P., Wilson, G. W., Tang, Y., & Scott, K. S. 2013, *MNRAS*, 436, 2535
- Juneau, S., Dickinson, M., Bournaud, F., et al. 2013, *ApJ*, 764, 176
- Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, *MNRAS*, 346, 1055
- Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, *ApJS*, 180, 330
- Kovač, K., Lilly, S. J., Cucciati, O., et al. 2010, *ApJ*, 708, 505
- Krawczyk, C. M., Richards, G. T., Mehta, S. S., et al. 2013, *ApJS*, 206, 4
- Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, *ApJS*, 154, 166
- Lang, D., Hogg, D. W., & Schlegel, D. J. 2014, arXiv:1410.7397
- Le Flocc'h, E., Aussel, H., Ilbert, O., et al. 2009, *ApJ*, 703, 222
- Lewis, A., & Bridle, S. 2002, *PhRvD*, 66, 103511
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
- Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, *MNRAS*, 425, 623
- Lusso, E., Comastri, A., Vignali, C., et al. 2011, *A&A*, 534, A110
- Lusso, E., Hennawi, J. F., Comastri, A., et al. 2013, *ApJ*, 777, 86
- Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, *MNRAS*, 449, 4204
- Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, *A&A*, 532, A90
- Madau, P. 1995, *ApJ*, 441, 18
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, 115, 2285
- Maiolino, R., Marconi, A., Salvati, M., et al. 2001, *A&A*, 365, 28
- Marconi, A., & Hunt, L. K. 2003, *ApJ*, 589, L21
- Marconi, A., Risaliti, G., Gilli, R., et al. 2004, *MNRAS*, 351, 169
- Mary Kathryn Cowles, B. P. C. 1996, *J. Am. Stat. Assoc.*, 91, 883
- McCracken, H. J., Capak, P., Salvato, M., et al. 2010, *ApJ*, 708, 202
- McCracken, H. J., Radovich, M., Iovino, A., et al. 2008, *yCat*, 2286, 0
- Meiksin, A. 2006, *MNRAS*, 365, 807
- Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, *MNRAS*, 437, 3550
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *JChPh*, 21, 1087
- Moller, P., & Jakobsen, P. 1990, *A&A*, 228, 299
- Murphy, E. J., Condon, J. J., Schinnerer, E., et al. 2011, *ApJ*, 737, 67
- Pirzkal, N., Rothberg, B., Nilsson, K. K., et al. 2012, *ApJ*, 748, 122
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A16
- Prescott, M. K. M., Impey, C. D., Cool, R. J., & Scoville, N. Z. 2006, *ApJ*, 644, 100
- Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E., & Rocca-Volmerange, B. 1984, *A&A*, 132, 389
- Prochaska, J. X., Madau, P., O'Meara, J. M., & Fumagalli, M. 2014, *MNRAS*, 438, 476
- Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006a, *ApJS*, 166, 470
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006b, *AJ*, 131, 2766
- Richstone, D. O., & Schmidt, M. 1980, *ApJ*, 235, 361
- Sajina, A., Scott, D., Dennefeld, M., et al. 2006, *MNRAS*, 369, 939
- Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
- Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T., & Matthews, K. 1989, *ApJ*, 347, 29
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Scott, J. E., Kriss, G. A., Brotherton, M., et al. 2004, *ApJ*, 615, 135
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Serra, P., Amblard, A., Temi, P., et al. 2011, *ApJ*, 740, 22
- Shang, Z., Brotherton, M. S., Wills, B. J., et al. 2011, *ApJS*, 196, 2
- Shull, J. M., Stevans, M., & Danforth, C. W. 2012, *ApJ*, 752, 162
- Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, *ApJ*, 509, 103
- Silva, L., Maiolino, R., & Granato, G. L. 2004, *MNRAS*, 355, 973
- Silverman, J. D., Lamareille, F., Maier, C., et al. 2009, *ApJ*, 696, 396
- Slone, O., & Netzer, H. 2012, *MNRAS*, 426, 656
- Soifer, B. T., Boehmer, L., Neugebauer, G., & Sanders, D. B. 1989, *AJ*, 98, 766
- Springel, V., Di Matteo, T., & Hernquist, L. 2005, *MNRAS*, 361, 776
- Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *ApJ*, 631, 163
- Stern, J., & Laor, A. 2012, *MNRAS*, 423, 600
- Stevans, M. L., Shull, J. M., Danforth, C. W., & Tilton, E. M. 2014, *ApJ*, 794, 75
- Stickel, M., Lemke, D., Klaas, U., et al. 2000, *A&A*, 359, 865
- Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, *ApJ*, 565, 773
- Tilton, E. M., Stevans, M. L., Shull, J. M., & Danforth, C. W. 2016, *ApJ*, 817, 56
- Trump, J. R., Impey, C. D., Elvis, M., et al. 2009, *ApJ*, 696, 1195
- Trump, J. R., Impey, C. D., Gabor, J. M., et al. 2007, in *ASP Conf. Ser. 373. The Central Engine of Active Galactic Nuclei*, ed. L. C. Ho & J.-W. Wang (San Francisco, CA: ASP), 726
- Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, *ApJ*, 598, 886
- Urry, C. M., & Padovani, P. 1995, *PASP*, 107, 803
- VanderPlas, J. 2014, arXiv:1411.5018
- Worseck, G., & Prochaska, J. X. 2011, *ApJ*, 728, 23
- Xue, Y. Q., Brandt, W. N., Luo, B., et al. 2010, *ApJ*, 720, 368
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
- Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, *ApJS*, 172, 468
- Zheng, W., Kriss, G. A., Telfer, R. C., Grimes, J. P., & Davidsen, A. F. 1997, *ApJ*, 475, 469