Dependence of Optical Active Galactic Nuclei Identification on Stellar Population Models

Yan-Ping Chen1, Ingyin Zaw1,2, and Glennys R. Farrar2
1 New York University Abu Dhabi, P.O. Box 129188, Abu Dhabi, United Arab Emirates; yc47@nyu.edu
2 Center for Cosmology and Particle Physics, Physics Department, New York University, New York, NY 10003, USA

Received 2017 December 6; revised 2018 May 23; accepted 2018 May 27; published 2018 July 5

Abstract

We have conducted a study to quantify the systematic differences resulting from using different stellar population models (SPM) in optical spectroscopic identification of type II active galactic nuclei (AGNs). We examined the different AGN detection fractions of 7069 nearby galaxies (z ≤ 0.09) with Sloan Digital sky Survey (SDSS) DR8 spectra when using the Bruzual & Charlot (BC03), Vazdekis et al. (MILES), and solar metallicity Maraston & Strömbäck (MS11_solar) SPM. The line fluxes obtained using BC03 and MS11_solar are publicly available from SDSS data releases. We find that the BC03 templates result in systematically higher BPT line ratios and consequently higher AGN fractions, and the MS11_solar templates result in systematically lower line ratios and AGN fractions compared with the MILES templates. Using MILES as the standard, BC03 results in 25% “false positives” and MS11_solar results in 22% “false negatives” when using the Kewley et al. boundary for AGN identification. The fraction of galaxies whose AGN identification changes for different templates is luminosity dependent, ranging from a few percent for $L_{\text{[O III]} 5007} \geq 10^{40}$ erg s$^{-1}$ and increasing to ~50% for $L_{\text{[O III]} 5007} \leq 10^{38}$ erg s$^{-1}$. These results suggest that template choice should be accounted for when using and comparing the AGN and emission line fluxes from different catalogs.

Key words: galaxies: active – galaxies: nuclei – techniques: spectroscopic

1. Introduction

Optical spectroscopy is a powerful tool to probe the physical properties of active galactic nuclei (AGNs). Broad-line (type I) AGNs are characterized by their broad hydrogen Balmer lines, with Full Width at Half Maximum values up to a few thousand km s$^{-1}$. Narrow-line (type II) AGNs, however, need to be distinguished from star-forming galaxies because they emit the same set of ionized forbidden lines, such as [N II] 6584 and [O III] 5007. The differences in the ratios of these lines to the narrow hydrogen Balmer lines (i.e., Hα, and Hβ) between AGNs and star-forming galaxies were first reported by Baldwin et al. (1981, hereafter, BPT), and further explored by others (e.g., Osterbrock & Pogge 1985; Veilleux & Osterbrock 1987; Kewley et al. 2001a, 2006; Kauffmann et al. 2003; Stasińska et al. 2006; Cid Fernandes et al. 2010, 2011). Kewley et al. (2001a) developed separation criteria based on theoretical modeling of star formation lines, while Kauffmann et al. (2003) defined empirical separation criteria based on Sloan Digital sky Survey (SDSS; York et al. 2000) spectra.

The spectra of galaxies containing AGNs show not only emission features but also stellar absorption lines and continuum emission from the host galaxies. Subtraction of the host galaxy contribution to the integrated light in previous works (e.g., Ho et al. 1997; Kewley et al. 2000, 2001b; Kauffmann et al. 2003) has been done either with a local polynomial fit or using templates (e.g., Bruzual & Charlot 2003). The former is easy to perform, does not rely on any models, and can be straightforwardly applied as a quick solution for the absorption and continuum components estimation, especially for objects with very strong emission features (e.g., those of Seyfert I AGNs). However, when the spectrum contains nonnegligible absorption features, the accuracy of the emission line fluxes can be affected. Weak emission line features, especially hidden in the absorption and continuum components, can easily be underestimated in a local polynomial continuum fit. In these cases, using stellar templates to fit the absorption and continuum gives a more accurate solution. Full-spectrum-fitting and subtraction of absorption and continuum components requires strong enough absorption features to match with stellar templates. Instead of full-spectrum-fitting, one can also use principle component analysis (e.g., Kewley et al. 2001a, 2006) using SDSS data. The AGN detection rate is known to depend on the details of data processing (Hao et al. 2005): redshift ranges, signal-to-noise cut of the data, and the boundary that separates type II AGNs from star-forming galaxies in BPT diagrams (i.e., the authors may have different AGN identification criteria).

In this work, we consider the impact of using different stellar population models (SPM) on AGN identification, which has not previously been quantified. The AGN identification in SDSS data releases (DRs) first used Bruzual & Charlot (2003, BC03), and currently Maraston & Strömbäck (2011) at solar metallicity only (MS11_solar), as templates for stellar component analysis. As more SPM become available, systematic studies of AGN identification’s dependence on the stellar model used become possible. This is important not only for finding the best template but also for understanding the merits and limitations of the models. In our study, we compare the results from BC03, MS11_solar, and Vazdekis et al. (2010, MILES) SPM.

In Section 2, we present both the sample for our study and our procedure for identifying narrow-line AGN candidates. In Section 3, we review the SPM used in this analysis. We also describe how we applied SPM to our data. Our results and the relative effects from varying the templates are presented in Section 4, where we also discuss the aspects of the stellar
templates that lead to different results. We present our conclusions in Section 5. The Appendix presents other properties we explored that do not cause major differences in type II AGN identification, namely, wavelength range, young stellar populations, Seyfert IIs versus LINERs, and data quality.

2. Data Sample and AGN Identification

This work is a part of our effort to construct an all-sky optical AGN catalog (I. Zaw et al. 2018, in preparation, hereafter ZCF18), based on optical spectra from the 2MASS Redshift Survey (2MRS; Huchra et al. 2012). The 2MRS was assembled from observations by the 2MRS team (with the FAST Spectrograph for the Tillinghast telescope at the Fred L. Whipple Observatory in the north and Cerro Tololo Inter-American Observatory in the south) and from other catalogs, including the SDSS DR 8, the 6dF Galaxy Survey, and the NASA Extragalactic Database. Among all of the subsamples, the SDSS subsample has the best signal-to-noise ratios (S/Ns) and is the only one where absolute flux calibration and telluric correction have been applied to the spectra. In addition, the line fluxes from SDSS DRs can be used to cross check our work. We therefore use the SDSS subsample for our study. The SDSS subsample consists of the spectra of 7069 galaxies with a redshift of $z \leq 0.09$. These spectra cover the wavelength range 3800–9200 Å, with a mean resolution of $R \sim 1800$–2000.

2.1. Type II AGN Candidate Identification

A detailed description of the emission line galaxy selection is given in our catalog paper (ZCF18). We briefly summarize the process here. After rebinning the data and model spectra to the same spectral resolution, and masking out the AGN identification emission line regions, a full-spectrum-fitting was performed on each SDSS object using pPXF (Cappellari & Emsellem 2004). In the fitting procedure, the model spectra are shifted to the redshift of the observed spectra and broadened to account for stellar velocity dispersion. Each fit consists of a linear combination of model spectra. The fits are required to yield a physical stellar velocity dispersion, $\sigma_{\text{fit}} < 1000$ km s$^{-1}$, in order to be considered acceptable. The fitting routine uses the error spectrum, provided with the data, to calculate a reduced $\chi^2$ value for each acceptable fit. The “best-fit” spectrum is the one that minimizes the reduced $\chi^2$. We limit the sample to those spectra whose reduced $\chi^2$ are less than 2.55 in the full-spectrum-fitting process, which keeps 99% of the spectra with successful fits.

The residual spectrum produced by subtracting the best-fit model from the data spectrum is used to analyze the emission line features. The galaxies with weak emission lines are most affected by the choice of stellar templates. Figure 1 shows an example where the H$\beta$ emission line is invisible in the spectrum by eye, but is identified by the template subtraction using MILES (Vazdekis et al. 2010) templates.

In this work, we use H$\alpha$, H$\beta$, [O III] 5007, and [N II] 6584 to identify galaxies as type II AGNs. We infer the flux for each of the emission lines from the residual spectrum, by fitting Gaussian profiles. The line fluxes are calculated under the fitted Gaussian profiles within 3$\sigma$ of the line peak, where $\sigma$ is the fitted Gaussian width of the line. The flux errors (i.e., line noises) are calculated as the sum in quadrature of the error spectra under the same wavelength range of the emission lines. An emission line is defined as one where the line flux divided by the flux error is greater than three (i.e., S/N $\geq$ 3) when using MILES templates. We define emission line galaxies as those where all four diagnostic lines (H$\beta$, [O III] 5007, H$\alpha$, and [N II] 6584) have S/N $\geq$ 3. There are 3350 emission line galaxies in our sample. We use the BPT line ratios [N II]/H$\alpha$ and [O III]/H$\beta$ and the Kewley et al. (2001a) criteria to separate star-forming galaxies and type II AGN candidates, i.e., those above the Kewley et al. (2001a) line are taken as type II AGNs.

![Figure 1](image_url)  
*Figure 1.* An example illustrating that template subtraction is necessary to identify weak emission lines. Top panels: data are marked by black lines, the best-fit models are marked by red lines, the horizontal blue lines show the 1σ error spectra, and the green dots represent the residuals in regions without emission lines. The weak Hβ emission line is absent before template subtraction. Bottom panels: residual spectra showing the emission line features after subtracting the best-fit (absorption) models.
those below the Kewley et al. (2001a) line are taken as star-forming galaxies.

3. Model Templates

SPM are constructed by integrating a group of stellar spectra, known as a stellar library, with weights given by the initial mass functions. The group of stars is assumed to share the same age and chemical components, but with a variety of stellar masses. Their integrated light forms a single stellar population (SSP) spectrum. The connection between SPM and individual stellar spectra are stellar parameters such as effective temperature $T_{\text{eff}}$, surface gravity $\log g$, and metallicity [Fe/H]. When the stellar populations are constructed, an interpolator is applied to generate the grid of stars used to integrate the light of the population; see Vazdekis et al. (2010), Conroy (2013) and references therein for details. The interpolator is based on the parameter space coverage of the underlying stellar library. When stars are limited to a certain part of the parameter space, the interpolator may be biased if the available data does not adequately span the parameter space. Each SPM consists of a set of SSP templates with a given age and metallicity.

The Bruzual & Charlot (2003) templates are widely used to subtract the stellar absorption components and the continuum, in analyses of emission line galaxies, because they cover a large spectral wavelength range and were developed earlier. In the last decade, with improvements in both data and theoretical modeling, more stellar libraries have been constructed and published (e.g., Prugniel & Soubsiran 2001; Gregg et al. 2006; Sánchez-Blázquez et al. 2006; Chen et al. 2014), and their corresponding SPM are available (e.g., Le Borgne et al. 2004; Vazdekis et al. 2010, 2016; Maraston & Strömbäck 2011). In this work, we compare the SPM of MILES, MS11$_{\text{solar}}$, MILES-based models, González Delgado et al. (2005, hereafter G05) and BC03. The SDSS fluxes for DR8 were based on host galaxy subtraction using BC03. More recently, Thomas et al. (2013) made public new fluxes for DR8 spectra using MS11$_{\text{solar}}$. The MILES model is based on observed stellar optical spectra (Sánchez-Blázquez et al. 2006); the BC03 model was constructed from a combination of theoretical and observed spectra. The MS11$_{\text{solar}}$ MILES-based model shares the same input stellar library as MILES model. The G05 model was constructed from theoretical stellar spectra and is the best available theoretical model. We use them when empirical data are incomplete, e.g., for young stellar populations. The model ingredients are described in Section 3.1. Because the MILES models are built from the MILES stellar library, currently the best empirical optical stellar library and widely used by several SPM (e.g., Conroy & Gunn 2010; Vazdekis et al. 2010, 2012; Maraston & Strömbäck 2011), we use the MILES templates as our standard.

3.1. Differences of SPM

SPM have several distinct properties. The parameter ranges for the models we compare in this work are given in Table 1. They vary in age and metallicity ranges, as well as in stellar libraries. We will investigate how each of these affects our results. The key ingredients of SPM that describe the individual spectra of stellar systems from different sources are as follows. BC03 mainly uses a combination of the empirical stellar library STELIB (Le Borgne et al. 2003) and a series of theoretical stellar libraries BaSeL (Lejeune et al. 1997, 1998; Westera et al. 2002). The BC03 SSPs are likely to be biased in the optical, having been built from the STELIB library that contains 249 stars, with very few stars at non-solar metallicities. The MILES model is based on the empirical MILES stellar library that covers a wide range of stellar parameter space (i.e., $T_{\text{eff}}$, $\log g$, and [Fe/H]), with four times more stars than in STELIB. The G05 model is based on a theoretical stellar library; it contains a wide age range, which is helpful in understanding the contribution from younger stellar components in host galaxies. The MILES-based MS11 model used the same stellar library as MILES, with a special procedure for integrating the light of thermally pulsing asymptotic giant branch (TP-AGB) stars. The MS11$_{\text{solar}}$ templates are the subset of MILES-based MS11 SPM that assume solar metallicity abundances.

We also note that the resolutions of different SPMs are different, but this does not have a significant impact on our results, as the SDSS spectra have lower or similar resolutions to the template libraries. In addition, the velocity dispersions in host galaxies (a few hundred km s$^{-1}$) are bigger than the differences in resolution (a few tens km s$^{-1}$) of the SPM whose resolution is determined by the stellar libraries used. When comparing the models with the observations, broadening the stellar models to fit the spectral features of the observations erases the resolution differences. Therefore, the resolution difference issue is not further discussed in this work.

3.2. Spectral Line Features of SPM

We first directly compare different stellar population templates from each SPM available with the same parameters (i.e., age, metallicity, and initial mass function), as the spectral line features are particularly important in the subtraction procedure to yield the type II AGN candidates. In Figure 2, we show the MILES and BC03 stellar population templates for representative ages of 63 Myr, 1 Gyr, and 12.5 Gyr at solar ([Fe/H] = 0.0) and sub-solar metallicity ([Fe/H] = −0.4). We
see that for a given population, these two sets of models agree best at older ages ($12.5 \text{ Gyr}$) with solar metallicity. At younger ages, i.e., $63 \text{ Myr}$ and $1 \text{ Gyr}$, inconsistencies are seen in both the Balmer lines and the overall shapes (i.e., the colors). At sub-solar metallicity, there are inconsistencies in the red, $\sim 6800-7400 \, \AA$, especially for older populations. Although we have normalized the models at $5500 \, \AA$, there are still some color discrepancies.

To better understand the similarities and differences of the models in the wavelength ranges of the four main type II AGN...
identification lines, we zoom in on the SSPs. In Figure 3(a), we see that at solar metallicity, these two sets of models agree with each other especially at older ages (12.5 Gyr). However, models show differences in the depth of Hα (Figure 3(b)) at younger ages, with a 7% difference observed for 1 Gyr models. In general, MILES templates show stronger absorption Balmer lines than those of BC03.

We show the MILES-based MS11_solar and compare them with MILES SSPs in Figures 4(a) and (b). Although both MS11_solar and MILES used the MILES stellar library as their optical input, we still observe remarkable line differences and continuum variation, due to the other inputs that comprise an SPM. The MS11 models tend to have slightly stronger Balmer absorption lines than MILES at the youngest and oldest ages, but they have weaker Balmer absorption lines than MILES at intermediate ages (1 Gyr).

4. Is Type II AGN Identification Template-dependent?

To check our method and the validity of our results, we compared our line fluxes with the values in the SDSS DR for the emission line galaxies identified by our MILES-based template subtraction. Note that we use the MILES SPM while SDSS DR8 uses BC03; therefore, some differences in emission line fluxes are expected. The comparison between the fluxes of diagnostic lines (Hβ, [O III] 5007, Hα and [N II] 6584) from our fits and those from SDSS shows a good linear correlation (see ZCF18). The Hβ line is the weakest in general, so the scatter is larger than in the other lines.

Although the line fluxes correlate well overall, AGN identification based on the line flux ratios [O III] 5007/Hβ and [N II] 6584/Hα may have significant differences. Because galaxies near the boundary are the most likely to shift categories (AGN to star-forming galaxies or vice versa), we begin to examine differences in line ratio by selecting a subsample of objects within 0.02 dex of the Kewley et al. (2001a) boundary using MILES stellar population templates for host galaxy subtraction. This subsample contains 145 emission line galaxies.

4.1. Type II AGN Identification Variations

Figure 5 shows the BPT diagram for this boundary sample, with MILES-based line ratios marked as blue triangles, BC03-based line ratios (Aihara et al. 2011) shown as gray crosses, and the MILES-based MS11_solar (Thomas et al. 2013) result shown as magenta triangles. The line ratios using BC03 are systematically higher than those using MILES. Evidently, more galaxies are classified as AGNs when the background galaxy subtraction is processed with BC03 templates, while the line ratios from using MS11_solar templates are systematically shifted into the composite region, i.e., between the Kewley et al. (2001a) and Kauffmann et al. (2003) boundaries. The systematic shift of line ratios from MS11_solar templates can also be seen in the histogram, especially in the line ratio of [O III] 5007/Hβ. MS11_solar templates result in higher [O III] 5007/Hβ ratios than MILES, while MS11_solar templates result in lower [O III] 5007/Hβ ratios than MILES. The line ratio of [N II] 6584/Hα remains in a similar range.

We then expanded our study to the full emission line sample defined by MILES template subtraction. The BPT diagram for these 3350 galaxies is shown in Figure 6. The narrow emission line galaxies are shown as grey crosses, and type II AGNs (MILES-based) are shown as blue triangles. In panels (b) and (c), we show the line ratios from fluxes when using BC03 and MS11_solar templates, respectively, for the same galaxies. The symbols reflect AGN identification based on MILES template subtraction. We see that BC03-based line ratios tend to move toward the AGN region, except for a few outliers. MS11_solar-based line ratios move in the opposite direction, and some of the MILES-based AGNs shift into the composite region.

In order to learn which type II AGNs are most sensitive to the choice of stellar templates, we examine two classes of misidentification, taking the result from MILES template-subtraction as the standard. The first class is type II AGNs identified by BC03 but not identified by MILES templates, i.e., “false positives;” the second class is type II AGNs identified by MILES templates but not identified by MS11_solar, i.e., “false negatives.” Because [O III] luminosity is an indicator of the AGN bolometric luminosity, we use it to determine whether the
misidentification rate depends on AGN activity. We show the \([\text{O III}]\) luminosity distribution of BC03 type II AGNs in Figure 7(a). The clear histogram shows the \([\text{O III}]\) luminosity of 562 type II AGNs identified by using BC03 templates, selected from the MILES emission line galaxy sample; the green histogram shows the false positives. We find that 25% of BC03 type II AGNs are false positives.\(^3\) The fainter the object is, the more likely it is to be misidentified as an AGN. We show the ratio of misidentification of type II AGN candidates as a function of \([\text{O III}]\) luminosity in Figure 7(b). The largest misidentification rate, \(\sim50\%\), is found for galaxies with \([\text{O III}]\) luminosity fainter than \(\sim10^{38}\) erg s\(^{-1}\).

\(^3\) False negatives from BC03 and false positives from MS11\textsubscript{solar} are negligible (less than 2%).

In Figure 7(c), type II AGNs identified by MILES templates are shown in the clear histogram. Those false negatives when using line ratios from MS11\textsubscript{solar} template subtraction are shown in the green histogram. A total misidentification rate of 22% is observed. We plot the misidentification rate as a function of \([\text{O III}]\) 5007 luminosity in Figure 7(d): objects with \([\text{O III}]\) 5007 luminosity fainter than \(10^{38}\) erg s\(^{-1}\) can have a misidentification rate as large as \(\sim50\%\).

Furthermore, we limit the 2MRS-SDSS sample to have S/N \(\geq 3\) for all four AGN diagnostic lines for all three sets of templates to avoid effects of low signal to noise, which results in 2300 remaining galaxies. As shown in Figure 8, there is still a clear systematic shift between BC03 and MILES line ratios. We calculate the misidentification fraction with the new sample and get a similar result: 21.5% of type II AGNs identified by using BC03 templates are not identified as AGNs when using
Similarly, 28.1% of MILES type II AGNs fall below the Kewley et al. (2001a) boundary when using MS11solar derived fluxes.

### 4.1.1. Misidentification of LINERs and Seyfert IIs

Low-ionization narrow emission-line regions (LINERs) were first defined by Heckman (1980) and are different from Seyfert II galaxies. Typical LINERs have spectra features dominated by lines arising from low-ionization states; their luminosities are similar to giant H II regions, and they are common. AGNs located above the Kewley et al. (2001a) boundary in the [O III]/Hβ–[N II]/Hα diagram can be either LINERs or Seyfert IIs. It is, therefore, pertinent to investigate if LINERs or Seyfert IIs dominate the misidentification rate. In addition, Hα equivalent width (EW) $< 3$ Å is proposed to be an indicator of LINERs whose emissions are from hot evolved stars (Cid Fernandes et al. 2010). A detailed discussion of this topic is presented in Appendix A. In summary, no significant difference is found in the misidentification rate between LINERs and Seyfert IIs. As none of the misidentified Seyfert IIs or LINERs have Hα EW $< 3$ Å, it is unlikely the misidentified LINERs in our sample are powered by hot evolved stars.

### 4.1.2. Dependence on Data Quality

Our study of the misidentified fraction as a function the overall quality of the spectrum, i.e., the continuum S/N ratio, is detailed in Appendix B. In summary, no strong correlation between the misidentified fraction and data quality is found.

### 4.2. Comparisons of Templates

Figure 9 illustrates the difference in the best fit using the MILES and BC03 SPM as templates, zooming in on the best-fit ratios around the line regions we use to identify type II AGN.
The best fit using the MILES model predicts stronger Balmer lines (i.e., $\text{H}_\alpha$ and $\text{H}_\beta$ lines) than that using BC03 templates. We show the flux variations between the results from BC03 and MILES template subtraction in Figure 10. The type II AGNs identified by MILES template subtraction are highlighted in red circles. The mean flux of $\text{H}_\beta$ from BC03 type II AGN is $\sim 26\%$ fainter than that from MILES templates, and the $\text{H}_\alpha$ line fit from BC03 is $\sim 6\%$ fainter. Most of the type II AGNs have stronger [O III] fluxes from BC03 than from MILES templates. The fluxes of [N II] from BC03 and MILES templates are consistent with each other. The inaccurate absorption lines of the BC03 model are also described in detail by other groups (e.g., González Delgado et al. 2005; Koleva et al. 2008).

The differences between MILES and MS11_solar are more surprising because they use the same input stellar library. This

**Figure 8.** BC03-based line ratios (gray crosses) for 565 galaxies with $S/N \geq 3$ above the Kewley et al. (2001a) boundary. MILES-based ratios (blue triangles) for the same galaxies systematically shift downwards.

**Figure 9.** An example showing a comparison of the best-fit models from the full-spectrum-fitting result using MILES and BC03 as templates. The “best-fit ratio” is the ratio between best fit derived by BC03 templates and the one derived by MILES templates.
difference could be due to the treatment of the TP-AGB stars affecting SSPs between 0.2 and 2 Gyr. However, we examined the stellar population components of the sample galaxies and found only 8% of them contain younger (age ≤ 2 Gyr) populations. In each of these galaxies, young stellar populations contribute less than 56% of the optical light. This means the systematic shift of line ratios when using MS11 solar is not dominated by the special treatment of TP-AGB stars.

Furthermore, we note that we used all of the available metallicities in the MILES model, and our best fits indicate that most (∼70%) of our galaxies favor metal-rich models, as shown in Figure 11. The green subsample highlights galaxies containing only one stellar population. While it is true that there is a degeneracy between age and metallicity when fitting only colors, as noted by Thomas et al. (2013), the degeneracy is lifted when simultaneously fitting both the absorption lines and the continuum (e.g., Reichardt et al. 2001). Therefore, the choice of metallicity range of the templates also plays a role in type II AGN identifications. A detailed comparison between the SPM by Maraston & Strömbäck (2011) and Vazdekis et al. (2010) is beyond the scope of this work and will not be addressed here.

4.3. Young Stellar Populations and Wavelength Range

Young stellar populations (≤63 Myr) are absent in all of the template families discussed above. However, as we show in Appendix C, the lack of young stellar populations does not significantly affect AGN identifications. We use G05 to explore young populations, as empirical young SPMs are not available, and discuss the consistency between the G05 model and the MILES model at 63 Myr. An expansion of the MILES model with the G05 young population models is presented and used for AGN identification around the Kewley et al. (2001a) border. We also examined differences due to wavelength ranges used in host galaxy subtraction in Appendix D. We find that young stellar populations (i.e., age ≤ 63 Myr) and the

Figure 10. The difference of line fluxes between BC03 and this work (using MILES). This work, BC03, and MS11 solar of the 2300 objects with S/N ≥ 3 of all four lines from both templates are shown by black crosses. Type II AGNs (i.e., galaxies with line ratios above the Kewley et al. (2001a) boundary in the [O III]/Hβ/[N II]/Hα diagram) based on MILES template subtraction are shown as red circles. Note that the line ratio scatters are large. We mark the mean values of the line ratios of the type II AGNs as blue lines in each panel.

Figure 11. The metallicity distribution of the major population components of SDSS DR8 galaxies from the full-spectrum-fitting result. The green histogram shows the ones that contain only one simple stellar population.

5. Conclusions

We have examined the differences in optical identification of type II AGNs in nearby (z ≤ 0.09) galaxies with SDSS DR8 spectra, resulting from host galaxy subtraction using MILES, BC03, and MS11 solar as stellar templates. We found that type II AGN identification is sensitive to the stellar template. Comparing the results of using the BC03 and MILES SPMs to subtract absorption lines in SDSS DR8 data, we determined that one-quarter of the sample is misidentified as type II AGNs by BC03 relative to MILES. Results using the MS11 solar templates show fewer galaxies identified as AGNs relative to MILES. We also find a 22% disagreement overall with the work of Thomas et al. (2013), which used MS11 solar for their host galaxy continuum and absorption subtraction. We traced the problem to the incomplete range of metallicities of the SPMs used in template fitting. The misidentification of both using BC03 (e.g., the work of MPA-JHU, SDSS DR8) and using MS11 solar (e.g., the work of Thomas et al. 2013) is greatest for objects with low [O III] luminosities and is up to 50% for [O III] 5007 luminosity fainter than 10^{38} erg s^{-1}. The MS11 solar used for the subtraction of the host galaxy contribution should be taken into account when using the emission line fluxes, or the AGN fractions, from a catalog especially if the results from different catalogs are compared.

We thank YuXiao Dai who helped us in checking the data and some useful discussions. We thank Jong-Hak Woo for the discussion about AGN luminosities and AGN identifications. We thank the referee for his/her valuable comments and suggestions. The research of GRF was supported by National Science Foundation grants NSF-PHY-1212538 and NSF-AST-1517319, and by the James Simons Foundation.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.
SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

Appendix A
Misidentification of LINERs and Seyfert IIs

Whether low-ionization nuclear emission-line regions (LINERs) are AGNs is a topic that has been debated in literature. Ho (1996, 2008) and Masegosa et al. (2011), for example, argued that a significant fraction of LINERs are low-luminosity AGNs. Other studies have suggested that LINERs are, instead, shock heated gas (Dopita & Sutherland 1995), starburst activity (Terlevich & Melnick 1985; Alonso-Herrero et al. 2000), or post-AGB stars (Singh et al. 2013). Cid Fernandes et al. (2010) suggest that Hα EW can differentiate between the different ionization mechanisms that lead to the overlap in the LINER region of traditional diagnostic diagrams. According to the bimodal distribution of Hα EW, Cid Fernandes et al. (2011) suggest that LINERs with Hα EW > 3 Å are likely to be true AGNs, while those with Hα EW < 3 Å have emissions from hot evolved stars.

In order to discern the nature of the LINERs in our sample using the Hα EW, we first separate the AGN identified from MILES-based line ratios into LINERs and Seyfert IIs using the Kewley et al. (2006) criteria. Figure 12 shows the misidentified LINERs and Seyfert IIs when using BC03 templates as a function of Hα EW. None of the galaxies are found to have Hα EW < 3 Å. Therefore, there are no LINERs powered by hot evolved stars, as defined by Cid Fernandes et al. (2010), in our misidentified sample. The misidentification instead mainly comes from line ratio variations due to template subtraction.

Furthermore, Figure 12 shows that no clear separation of Hα EW distribution is observed between the misidentified LINERs and Seyfert II galaxies. The line ratios based on the MILES template-subtraction is shown in Figure 13(a). Galaxies are limited to the sample with S/N ≥ 3 for all four lines in all three templates. Seyfert IIs are shown in purple and LINERs are shown in dark green. We also show the line ratios based on BC03 in panel (b) and line ratios based on MS11solar in panel (c) for the same galaxies. As clearly shown in the plots, line ratios derived from the different template-subtractions show systematic shifts. BC03-based line ratios shift toward the AGN/Seyfert II regions, while the MS11solar-based line ratios show the opposite trend.

Of the Seyfert IIs identified using BC03, 20.7% fall below the Kewley et al. (2001a) boundary using MILES-based line ratios. Similarly, the false positive rate of the BC03 LINERs is 14.6%. Our investigation of MS11solar-based line ratios shows a 26.0% false negative rate for Seyfert IIs and a 27.7% false negative rate for LINERs. Roughly half of the misidentified galaxies are (were) Seyfert IIs and half are (were) LINERs.
Appendix B

Dependence on Data Quality

We examine the misidentified AGN fraction against the overall data quality of the spectrum, i.e., the continuum S/N ratio. Because S/N varies at different wavelengths, we choose the continuum S/N near the Hβ line (hereafter S/N(Hβ)) for this purpose, as Hβ is the weakest of the four type II AGN identification lines. Figure 14(a) shows the distribution of S/N(Hβ) for type II AGN. The green histogram highlights the type II AGN identified when using BC03 templates, but not as type II AGN, when using MILES templates. The misidentification rate shows a weak dependence of S/N(Hβ) as seen in Figure 14(b). Exploration of the dependence between [O III] 5007 luminosity and S/N(Hβ) is shown in Figure 14(c). Misidentified type II AGN are shown as filled red triangles. No strong correlation is found.

We show the result from MS11_solar templates in Figure 15. The distribution of S/N(Hβ) for type II AGNs identified from MILES templates but not from MS11_solar templates is shown in the green histogram. The misidentification rate as a function of S/N(Hβ) is shown in Figure 15(b). The result is similar to the comparison between MILES and BC03, except for the first S/N bin that contains low statistics. We also explore the dependence between [O III] 5007 luminosity and S/N(Hβ). As shown in Figure 15(c), misidentified type II AGNs are shown as filled red triangles. No strong correlation is found. The misidentification of type II AGNs is, therefore, not due to data quality.

Figure 13. Panel (a): line ratios based on MILES template-subtraction to show the distribution of star-forming galaxies, Seyferts, and LINERs. S/N ≥ 3 were set up for all four lines from all three templates. The Kewley et al. (2006) boundary (the blue dashed line) is adopted to distinguish Seyferts and LINERs, where Seyferts are shown as purple circles, and LINERs are shown as green circles. Panel (b): line ratios based on BC03 for the same galaxies in panel (a). Panel (c): line ratios based on MS11_solar for the same galaxies in panel (a). The colors for panels (b) and (c) are based on MILES classifications.
Figure 14. BC03-based misidentification as a function of data quality near H/β line. Panel (a): the clear histogram shows all type II AGNs identified by using BC03 templates. The misidentified type II AGNs are shown in filled green histogram. Panel (b): the misidentification rate as a function of continuum S/N near H/β line. Panel (c): the luminosity of [O III] 5007 of type II AGNs from BC03 template subtraction as a function of continuum S/N near H/β line are shown as black diamonds. The misidentified type II AGNs are marked by filled red triangles.
Appendix C
Effects of Young Stellar Populations

C.1. Young Stellar Populations

Comparing parameters of the BC03 and MILES models in Table 1, we note that BC03 contains younger stellar populations than MILES. Young stellar populations are hard to model due to the lack of empirical observations. BC03 models have young stellar populations with a corrected continuum, but their lines have not been corrected (Bruzual & Charlot 2003, Section 2.2.3), especially at non-solar metallicities (González Delgado et al. 2005). We instead resort to theoretical young stellar populations. We expand the MILES model by adding young theoretical stellar population templates from G05 to make up for the fact that there are not many empirical stellar libraries covering that parameter space. As pointed out by Charlot & Fall (2000), stellar populations younger than \( \sim 3–4 \) Myr do not contribute to the observed spectra, because their absorption features are hidden behind their optically thick H II cloud. We therefore have confidence that using the models by González Delgado et al. (2005) at a youngest age of 4 Myr is adequate. To check the consistency between the G05 theoretical model and MILES model, we compare the spectra of their populations at the common age of 63 Myr, the youngest stellar population in the MILES model, at metallicities of \([\text{Fe/H}] = -0.4\) and \([\text{Fe/H}] = 0.0\) (solar metallicity). As shown in Figure 16, the G05 and MILES models are generally consistent with each other even at the boundary of the MILES model, with less than 10% deviation in their residual spectra.

Figure 15. MS11_{solar}-based misidentification as a function of data quality near H\(\beta\) line. Panel (a): the clear histogram shows all type II AGNs identified by using MS11_{solar} templates. The misidentified type II AGNs are shown in the filled green histogram. Panel (b): the misidentification rate as a function of continuum S/N near H\(\beta\) line. Panel (c): the luminosity of \([\text{O III}] 5007\) of type II AGNs from MS11_{solar} template subtraction as a function of continuum S/N near H\(\beta\) line are shown as black diamonds. The misidentified type II AGNs are marked by filled red triangles.

Figure 16. G05 templates of 63 Myr at two metallicities compared with the MILES model. G05 templates were smoothed to the same resolution as MILES library of FWHM = 2.5 Å. The residual spectra are the difference between these two sets of models.
The consistency discussed above gives us confidence that mixing theoretical and empirical models does not introduce systematic effects. Because the wavelength range in González Delgado et al. (2005) is slightly shorter than in MILES, we truncate the MILES model to the common wavelength range \( \lambda \lambda 3500-7000 \) Å. All SPM younger than 63 Myr from González Delgado et al. (2005) were broadened to the same resolution as MILES.

Following our initial strategy, we limit the sample within 0.02 dex around the Kewley et al. (2001a) boundary in the BPT diagram to investigate the line ratio variations. We fit the SDSS spectra using the young-population-extended MILES templates. We found that only three objects out of the 145 boundary galaxies contain young population components, with a contribution less than four percent in each of the cases. Fits for the other 142 galaxies give almost exactly the same result as using only MILES templates in the same wavelength coverage, as shown in Figure 17. In fact, even those three objects show only small shifts in the line ratios, \( \sim 0.03 \) in \([\text{O III}]/\text{H}\beta\) and \( \sim 0.002 \) in \([\text{N II}]/\text{H}\alpha\), far smaller than the typical line ratio errors. As shown by the olive diamonds (MILES + G05 young populations) overlapped with the red dots (MILES only, wavelength truncated to the same range as MILES + G05), there are no significant changes resulting from adding younger templates.

**Appendix D**

**Wavelength Range Dependence**

We also considered whether the wavelength range used for the spectrum fitting and absorption and continuum component subtraction may play a role in AGN identification. We performed this test using both the BC03 templates and MILES templates.

As mentioned above, we expanded the MILES models with young stellar populations with G05 with truncated wavelength 3500–7000 Å, as the red wavelength limit of G05 models is 7000 Å. The line ratio variation from wavelength truncation of MILES templates is shown in Figure 17. Compared to the line ratios derived from the full wavelength range of MILES, \( \lambda \lambda 3800–7500 \) Å, the result with truncated MILES templates shows scattered line ratios around the Kewley et al. (2001a) boundary but no strong systematic effects. Figure 17 illustrates the distribution of the line ratios: by truncating only 400–500 Å in the red, the line ratios scatter more around the Kewley et al. (2001a) boundary, sharing the similar ranges of \([\text{O III}]/\text{H}\beta\) and \([\text{N II}]/\text{H}\alpha\). We compare the line ratio differences in Figure 18. It shows that the center of \([\text{N II}]/\text{H}\alpha\) is shifted by 0.04 dex, and the center of \([\text{O III}]/\text{H}\beta\) is shifted by \(-0.05\) dex. The line ratio offsets of both \([\text{N II}]/\text{H}\alpha\) and \([\text{O III}]/\text{H}\beta\) are consistent with zero within the errors.

We changed the fitting wavelength range to \( \lambda \lambda 3800–7500 \) Å in testing the BC03 templates. The results are shown in Figure 19 to illustrate the difference from the original fitting wavelength of \( \lambda \lambda 3800–9200 \) Å. The gray histograms show the distribution of SDSS DR8 line ratios, and the golden histograms show the line ratios derived by truncated spectrum-fitting using BC03. The line ratios from the truncated wavelength fit still scatter around the same region in the BPT diagram as the result from wider wavelength range.

A detailed systematic analysis on the line ratios of truncated BC03 templates is shown in Figure 20. A shift of 0.08 dex is observed in \([\text{O III}]/\text{H}\beta\), and a shift of \(-0.01\) dex is observed in

**Figure 17.** The same galaxies based on MILES template subtraction within 0.02 dex around the Kewley et al. (2001a) boundary selected as in Figure 5, with the addition of the line ratios derived from young-stellar-population-extended MILES templates (red dots) and wavelength-truncated MILES templates (olive diamonds). The agreement of red dots and olive diamonds shows that adding young stellar population templates does not change the line ratios. The line ratios derived from truncated templates cause a larger scatter around the Kewley et al. (2001a) boundary, The distributions of line ratios are historgramed on the side of the axes. The colors of the histograms are the same as for the BPT diagram.
Figure 18. Distribution of line ratio differences between the fits with MILES model in wavelength range of 3800–7500 Å and 3800–7000 Å. The left panel shows the difference in the [N II]/Hα ratio; the right panel shows that difference in the [O III]/Hβ ratio.

Figure 19. BC03-based line ratios for 145 objects near the Kewley et al. (2001a) boundary as defined in Figure 5. The gray crosses are directly from the SDSS DR8 archive, the purple triangles are from this work, and the golden dots are from the wavelength-truncated BC03 templates. The distributions of line ratios are histogramed on the side of the axes. The colors are the same as those of the line ratios.
Again, the line ratio offsets of both [N II]/Hα and [O III]/Hβ are consistent with zero within the errors.

ORCID iDs
Yan-Ping Chen https://orcid.org/0000-0001-8821-0309
Ingyin Zaw https://orcid.org/0000-0002-5208-1426
Glennys R. Farrar https://orcid.org/0000-0003-2417-5975

References
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 195, 26
Allen, J. T., Hewett, P. C., Richardson, C. T., Ferland, G. J., & Baldwin, J. A. 2013, MNRAS, 430, 3510
Alonso-Herrero, A., Rieke, M. J., Rieke, G. H., & Shields, J. C. 2000, ApJ, 530, 688
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
Chen, Y.-P., Trager, S. C., Peletier, R. F., et al. 2014, A&A, 565, A117
Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, MNRAS, 413, 1687
Cid Fernandes, R., Stasińska, G., Schlickmann, M. S., et al. 2010, MNRAS, 403, 1036
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833
Dopita, M. A., & Sutherland, R. S. 1995, ApJ, 455, 468
González Delgado, R. M., Cerviño, M., Martins, L. P., Leitherer, C., & Hauschildt, P. H. 2005, MNRAS, 357, 945
Greene, J. E., & Ho, L. C. 2007, ApJ, 667, 131
Gregg, M. D., Silva, D., Rayner, J., et al. 2006, in The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode, ed. A. M. Koekemoer, P. Goudfrooij, & L. L. Dressel (Washington, DC: NASA), 209
Hao, L., Strauss, M. A., Tremonti, C. A., et al. 2005, AJ, 129, 1783
Heckman, T. M. 1980, A&A, 87, 152
Ho, L. C. 1996, in ASP Conf. Ser. 103, In The Physics of LINERs in View of Recent Observations, ed. M. Eracleous et al. (San Francisco, CA: ASP), 103
Ho, L. C. 2008, ARA&A, 46, 475
Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, ApJS, 112, 315
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001a, ApJ, 556, 121
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Kewley, L. J., Heisler, C. A., Dopita, M. A., et al. 2000, ApJ, 530, 704
Kewley, L. J., Heisler, C. A., Dopita, M. A., & Lumsden, S. 2001b, ApJS, 132, 37
Koleva, M., Prugniel, P., Ocvirk, P., Le Borgne, D., & Soubiran, C. 2008, MNRAS, 385, 1998
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, A&A, 425, 881
Le Borgne, J.-F., Bruzual, G., Pelíó, R., et al. 2003, A&A, 402, 433
Lejeune, Th., Cuisinier, F., & Buser, R. 1997, A&A, 125, 229
Lejeune, Th., Cuisinier, F., & Buser, R. 1998, A&A, 330, 65
Maraston, C., & Strömbäck, G. 2011, MNRAS, 418, 2785
Maségosa, J., Márquez, I., Ramírez, A., & González-Martín, O. 2011, A&A, 527A, 23
Miller, C. J., Nichol, R. C., Gómez, P. L., Hopkins, A. M., & Bernardi, M. 2003, ApJ, 597, 142
Osterbrock, D. E., & Pettigrew, R. W. 1985, ApJ, 297, 166
Prugniel, Ph., & Soubiran, C. 2001, A&A, 369, 1048
Reichardt, C., Jimenez, R., & Heavens, A. F. 2001, MNRAS, 327, 849
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Singh, R., van de Ven, G., Jahnke, K., et al. 2013, A&A, 558A, 43
Stasińska, G., Cid Fernandes, R., Mateus, A., Sodrè, L., & Asari, N. V. 2006, MNRAS, 371, 972
Terlevich, R., & Melnick, J. 1985, MNRAS, 213, 841
Thomas, D., Steele, O., Maraston, C., et al. 2013, MNRAS, 431, 1383
Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, MNRAS, 463, 3409
Vazdekis, A., Ricciardelli, E., Canara, A. J., et al. 2012, MNRAS, 424, 157
Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, MNRAS, 404, 1639
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295
Westera, P., Lejeune, T., Buser, R., Cuisinier, F., & Bruzual, G. 2002, A&A, 381, 524
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

Figure 20. Distribution of line ratio differences between the fits with BC03 templates in wavelength range of 3800–9200 Å and 3800–7400 Å. The left panel shows the difference of the line ratio [N II]/Hα; the right panel shows that difference of line ratio [O III]/Hβ.