Primordial $^4$He abundance refinement using sample of SDSS DR15 galaxies

O A Kurichin$^1$, A V Ivanchik$^1$ and V V Klimenko$^1$

$^1$ Ioffe Institute, Polytechnicheskaya, 26, 194021, Saint-Petersburg, Russia
E-mail: o.chinkuir@gmail.com

Abstract. We present spectroscopic data on metal-deficient HII regions from dwarf galaxies of the Local Volume on the basis of Sloan Digital Sky Survey (SDSS) Data Release 15 observations. Using the data on the intensities of the optical emission lines of the HII regions, one can determine their chemical composition, as well as other physical properties, such as temperature and electron density. Oxygen abundance (and thus the metallicity) of the source is determined using direct method by using strong [OIII] lines. Here we present only sources with oxygen abundance below $12 + \text{log}O/H = 8.5$. These data is used to determine primordial helium abundance, which value should be determined with as little errors as possibles. The main way to reduce errors is to use the spectroscopic data of the highest quality possible. On the other hand one can statistically decrease evaluation errors by using large dataset. Main focus of this paper is studying if we can decrease errors in $Y_p$ determination by combining HeBCD sample with SDSS spectra of lower quality.

1. Introduction

Studying elemental abundances in low-metal star-forming dwarf galaxies is important because their interstellar medium is said to be similar with the Early Universe one. Thus these galaxies and namely HII regions in them could be used to study key features of processes taking place at the beginning of Universe. On the other hand they could provide physically correct constraints on the numerical models of star chemical evolution. But in this paper we will focus only application of this data to primordial helium abundance determination.

Helium is the second most abundant element in the Universe. It was produced in first hundred seconds after Big Bang and it abundance increased due to subsequent stellar nucleosynthesis. Primordial helium abundance is called $Y_p$ and its value is about 0.25 related to hydrogen abundance. Importance of accurate $Y_p$ determination lines in to two main features. Firstly, it allows to test correctness of SBBN numerical models, which calculate the $Y_p$ as function of the baryon-to-photon ratio $\eta$. Secondly, helium abundance is highly sensitive to non-standart physics in primordial nucleosynthesis due to its sensibility to neutron lifetime $\tau$ and number of neutrino species $\nu$. It means that the $Y_p$ should be determined as accurate as possible. Classical way to determine this value is to calculate helium abundance in various sources then extrapolate obtained data to zero metallicity. It means that the way to reduce the $Y_p$ errors lies in reduction of $Y$ errors, which mainly can be done by obtaining more high quality spectrum. Upper bound in metallicity range is explained by fact, that there isn’t reliable star chemical evolution model, so it is unknown how well linear regression model will work with high-metal sources.
To obtain required level of accuracy one must use high quality spectra and big sample of objects to maximally reduce statistical and systematic uncertainties. Such sample of a total of 86 low-metallicity extragalactic HII regions in 77 galaxies had been gathered and introduced in [3] and is called HeBCD dataset. The problem with this sample is that it was collected from various papers spread through almost 15 years, and now we have better instruments for observations. On the other hand not all of named 86 sources are used in final analyses [2] so finale sample is not that good. Here we focus on studying effects of enlarging HeBCD sample with some "lower-quality" SDSS objects.

The paper is organized as follows. In Sect. 2 we describe our methodology to processing spectroscopic data, results of processing are presented in Sect. 3. Application of this new data to $Y_p$ determination and conclusions are summerized in Sect. 4.

2. Spectra processing procedure overview
Most common way to measure emission-line flux of such sources is to use IRAF SPLOT routine (see e.g. [4]). This time-tested routine have proven its robustness and accuracy, but IRAF is 35 years old legacy code and institutional support for IRAF and its usage is going away quickly, so it makes more sense to start new project using more up-to-date soft. We have chosen soft for processing one-dimensional spectra which have been developed for quasar spectroscopy.

We adopted its absorption lines processing features for measuring emission lines intensities and equivalent widths (EW). Since we are not interested in absolute flux value of single emission line our proposed routine works as follows:

- Upload spectra and construct its continuum. Because HII regions’ spectra have easily distinguishable one it could be fitted automatically using built-in procedures and than mistakes could be corrected manually.
- Specific emission lines are fitted by Gaussian functions, which best-fit parameters are determined by MCMC minimization. Then area under Gaussian curve ($A(\lambda)$) is calculated as a surrogate for line flux.
- To calculate flux ration of a line with wavelength $\lambda$ to $H\beta$ following relation is used:
  \[ \frac{I(\lambda)}{I(H\beta)} = \frac{A(\lambda)}{A(H\beta)} \]
- Equivalent width of line is determined by normalization of $A(\lambda)$ to continuum at line’s wavelength.

In order to test validity of this procedure we generated a synthetic spectrum using CLOUDY photo-ionization code. To make CLOUDY output spectra we convolved it with gaussian hardware function with resolution taken to be 2000. After that errors have been applied with signal-to-noise ratio(SNR) being equal to 10 which is in mid-range of typical SNR values for HII region in SDSS catalog and gaussian noise was added. This spectrum is presented on figure 1. Primary task is to determine the current helium abundance $Y$ in object. To do that we use method presented in [7]. It directly involves 4 Balmer series lines ($H\alpha - H\delta$) and 7 HeI lines ($\lambda$3889, $\lambda$4026, $\lambda$4471, $\lambda$5876, $\lambda$6678, $\lambda$7065, $\lambda$10830) plus one needs to know oxygen abundance in a source and temperature of OIII ion. To calculate these quantities 4 oxygen lines ($[OIII] : \lambda$4363, $\lambda$4959, $\lambda$5007 and $[OII] \lambda$3727) and 2 $[SII]$ lines ($\lambda$6717, $\lambda$6731) are needed, therefore we will measure intensities ratio only for this specific lines. Processing results and its comparison with CLOUDY calculation of named intensities are presented in table 1.

As seen from table 1 emission lines intensities ratios’ central values are retrieved with accuracy $\sim 1\%$ compared to generated ones with obtained uncertainties less than 2%. Thereby our proposed method of spectra processing could be used for real object spectrum analyses.
Figure 1. CLOUDY synthetic spectrum with marked lines for $Y$ determination

Table 1. Results of synthetical spectrum processing.

| Line            | $100 \times F(\lambda)/F(H\beta)^a$ | $100 \times F(\lambda)/F(H\beta)^b$ |
|-----------------|--------------------------------------|--------------------------------------|
| 3727.00 [O II]  | 68.90 ± 0.63                        | 37.50                                 |
| 3889.00 He I + H8 | 19.26 ± 0.24                        | 10.751                                |
| 4026.19 He I    | 2.30 ± 0.09                         | 2.37                                  |
| 4101.74 Hδ      | 26.52 ± 0.30                        | 26.25                                 |
| 4340.47 Hγ      | 48.18 ± 0.48                        | 47.42                                 |
| 4363.21 [O III] | 9.63 ± 0.14                         | 9.61                                  |
| 4471.48 He I    | 4.99 ± 0.11                         | 5.01                                  |
| 4861.33 Hβ      | 100 ± 0.90                          | 100                                   |
| 4958.92 [O III] | 237.26 ± 2.17                       | 236.72                                |
| 5006.80 [O III] | 725.27 ± 6.87                       | 706.27                                |
| 5875.60 He I    | 14.11 ± 0.07                        | 13.82                                 |
| 6562.80 Hα      | 281.84 ± 2.51                       | 276.56                                |
| 6678.10 He I    | 3.55 ± 0.08                         | 3.60                                  |
| 6716.40 [S II]  | 2.66 ± 0.07                         | 2.98                                  |
| 6730.80 [S II]  | 3.09 ± 0.07                         | 3.32                                  |
| 7065.30 He I    | 6.94 ± 0.11                         | 6.94                                  |
| 10830.34 He I   | 97.03 ± 0.95                        | 95.39                                 |

$^a$ CLOUDY spectra processing  
$^b$ direct CLOUDY calculation
3. Application to real objects

Data obtained from SDSS DR15 catalog [1] offers spectra with lower SNR compared to data observed with other telescopes [3, 4]. Lower SNR leads to inability to reliably detect some weak He I lines and increment in intensity uncertainties. Using list of extreme metal deficient HII regions from [5] and several other HII regions with mid and high metallicity we studied dependence of \( Y_p \) estimation on "low-quality" SDSS data incorporated with "high-quality" HeBCD data [3, 7]. List of used sources is presented in table 2.

| Name          | R.A.(J2000.0) | Dec.(J2000.0) | Redshift |
|---------------|---------------|---------------|----------|
| J0949+0036    | 09:49:54.1    | +00:36:58.64  | 0.006317 |
| J1045+0104    | 10:45:54.8    | +01:04:05.84  | 0.026199 |
| J1152-0040    | 11:52:47.5    | -00:40:07.66  | 0.004619 |
| J1226-0115    | 12:26:22.7    | -01:15:12.33  | 0.006648 |
| J0115-0051    | 01:15:34.4    | -00:51:46.03  | 0.005585 |
| J0814+4904    | 08:14:47.5    | +49:04:00.74  | 0.001876 |
| J0147+1356    | 01:47:07.0    | +13:56:29.29  | 0.056623 |
| J0015+0104    | 00:15:20.7    | +01:04:36.99  | 0.006858 |
| J0122+0048    | 01:22:41.6    | +00:48:42.00  | 0.057315 |
| J0137+1810    | 01:37:54.44   | 18:10:35.98   | 0.065934 |
| J0141+2124    | 01:41:33.22   | 21:24:50.33   | 0.163843 |
| J1304-0333    | 13:04:32.3    | -03:33:22.12  | 0.004502 |
| J0122+0057    | 01:22:13.9    | +00:57:31.43  | 0.007371 |

Extracting flux ratios same as in section 2 we calculated metallicity of a source by same method as [6] and helium abundance by method from [7]. Calculated physical parameters of each source are presented in table 3. This results could be easily incorporated with obtained \( Y \) values for various sources from HeBCD sample from [7]. Regression model with only HeBCD point and combined HeBCD and SDSS points are presented on figure 3. Linear regression have been carried our using MCMC method, regression results are presented on figure 2 and marginalized posterior probability distributions are presented on figure 3.

4. Conclusion

Using sources available in SDSS DR15 have both pros and cons. As seen from figure 3 for metal deficient sources (with \( O/H < 3 \times 10^{-5} \)) usage of SDSS spectra is not reasonable because low SNR leads to higher scatter of points near zero metallicity, which statistically increases estimated \( Y_p \) value. On the other hand for sources with \( O/H \) higher than \( 15 \times 10^{-5} \) errors in intensities for SDSS objects are comparable or even less than for same objects from HeBCD sample.

From figure 3 it is seen that HeBCD+SDSS dataset marginalized probability distributions are narrower and have less uncertainties compared to pure HeBCD dataset. Estimated result for combined dataset \( Y_p = 0.2468 \pm 0.0032, \ dY = 70 \pm 39 \) is somewhat higher compared to [7] result \( Y_p = 0.2449 \pm 0.0040, \ dY = 79 \pm 43 \), but estimated errors decreased by \( \sim 25\% \). It should be pointed that analyses carried out for SDSS points did not include the near-infrared line \( He\lambda 10830 \) because of missing data. But even in spite of this SDSS points juxtapose regression model quite well with good dispersion, so potentially adding SDSS points can further reduce uncertainty of \( Y_p \). Our estimation is that to reduce error of [7] evaluation one needs to insert \( \sim 50 \) SDSS point in linear regression data set alongside with HeBCD points.

In addition to mentioned above, according to [8] linear regression model works finely only at metallicities beneath \( 15 \times 10^{-5} \). Adding large amount of source above this threshold could be used to pick up model which fits chemical evolution better than linear regression. All this facts combined can help to reduce uncertainty of \( Y_p \) determination dramatically.
Table 3. Physical parameters of SDSS sources.

| Name            | T(OIII), 10^4 K | N_e, cm^{-3} | O/H ×10^5 | Y       |
|-----------------|-----------------|--------------|-----------|---------|
| J0015+0104      | 2.01 ± 0.82     | 339 ± 312    | 0.49 ± 0.40 | 0.2298 ± 0.0374 |
| J0122+0048      | 2.13 ± 0.22     | 251 ± 469    | 1.51 ± 0.32 | 0.2476 ± 0.0292 |
| J0137+1810      | 1.58 ± 0.16     | 554 ± 477    | 4.63 ± 1.85 | 0.2613 ± 0.0331 |
| J0141+2124      | 1.79 ± 0.17     | 852 ± 2679   | 3.60 ± 0.78 | 0.2851 ± 0.0484 |
| J0949+0036      | 1.10 ± 0.03     | 20 ± 19      | 10.90 ± 0.81 | 0.2462 ± 0.0057 |
| J1045+0104      | 1.15 ± 0.01     | 11 ± 1       | 10.73 ± 0.38 | 0.2654 ± 0.0082 |
| J1152-0040      | 1.51 ± 0.01     | 171 ± 43     | 5.90 ± 0.20  | 0.2518 ± 0.0128 |
| J1226-0115      | 1.51 ± 0.02     | 182 ± 42     | 5.91 ± 0.21  | 0.2509 ± 0.0115 |
| J0115-0051      | 0.93± ± 0.05    | 10 ± 1       | 13.59 ± 2.62 | 0.2584 ± 0.0046 |
| J0814+4904      | 1.35 ± 0.01     | 71 ± 21      | 10.86 ± 0.28 | 0.2428 ± 0.0053 |
| J0147+1356      | 1.42 ± 0.01     | 251± ± 44    | 8.60 ± 0.25  | 0.2648 ± 0.0131 |
| J0542-0121      | 1.04 ± 0.03     | 305± ± 274   | 22.3± ± 1.7  | 0.2710 ± 0.0090 |
| J0122+0057      | 0.99 ± 0.11     | 177± ± 188   | 22.0± ± 2.5  | 0.2678 ± 0.0132 |

Figure 2. Linear regression for only HeBCD points from [7] (blue) and incorporated HeBCD+SDSS points (red). Y_p value for HeBCD sample is shown as a light green dot, Y_p for HeBCD+SDSS sample is dark green dot.
Figure 3. The marginalized posterior probability distributions from the MCMC analysis.

Acknowledgments
This work was supported by Russian Science Foundation 18-12-00301.

We would like to thank S. A. Balashev for providing soft for spectra processing and expert advice on spectroscopy.

References
[1] Aguado et al 2019 ApJS 240 23
[2] Kurichin O A and Ivanchik A V 2018 JoP:CS 1135 012006
[3] Izotov Y I and Thuan T X and Stasińska G 2007 ApJ 662 15-38
[4] Izotov Y I and Thuan T X 2004 ApJ 602 200-30
[5] Guseva N G and Izotov Y I and Fricke K J and Henkel C 2017 AAp 599 A65
[6] Izotov Y I and Stasińska G and Meynet G and Guseva N G and Thuan T X 2006 AAp 448 955-70
[7] Aver E and Olive K A and Skillman E D 2015 JCAP 7 11
[8] Aver E and Olive K A and Skillman E D 2010 JCAP 5 003