Systematic effects and self-consistency in the primordial helium abundance determination

Oleg A. Kurichin, Alexandre V. Ivanchik

1 Ioffe Institute, Polytechnicheskaya, 26, 194021, Saint-Petersburg, Russia

E-mail: o.chinkuir@gmail.com

Abstract. Observations of light element abundances (\(^4\)He, D, \(^7\)Li) provide a powerful and independent test of the Standard Model of Big Bang Nucleosynthesis. The \(^4\)He abundance is measured most accurately among the light elements. This value is determined from observations of spectral lines from small-scale extragalactic sources (HII regions) in the optical and near infrared ranges. The intensity of the forbidden lines like [OIII] and [SII] allows to obtain the physical characteristics of the source, such as temperature and metallicity. Intensities of the \(^4\)He lines are used to determine the abundance of this element in the object. However, there are various systematic effects that should be taken into account for precise definition of this physical quantity. These include (1) interstellar reddening, (2) emissivities of \(^4\)He in different wavelengths, (3) underlying absorption for H and He lines, (4) temperature and (5) ionization structure HII regions, (6) collisional and fluorescent excitation of H and He lines. Incorrect accounting of these effects can lead to overestimation or underestimation of the primordial helium abundance \(Y_p\) by several percent.

At the moment, there are two independent groups of researchers dealing with this problem. The results for the calculation using the same data sample for these groups are as follows: \(Y_p = 0.2551 \pm 0.0022\) (Izotov, Thuan, Guseva, 2014) and \(Y_p = 0.2449 \pm 0.0040\) (Aver, Olive, Skillman, 2015). The last result is consistent with the result \(Y_p = 0.2467 \pm 0.0006\) based on the study of the CMB anisotropy (P.A.R. Ade et al., 2015). However, the values of \(Y_p\) obtained by these groups do not coincide even within the margin of error, and at the moment there is no clear answer to the question of which of these two groups got correct result. In the paper, differences in the methods of processing data and obtaining the quantity of the primordial helium abundance are investigated, in particular, how the groups take into account the named systemic effects. An error analysis and comparative tests are performed.

1. Introduction

Helium is the second most abundant element in the Universe. It was produced in first hundred seconds after Big Bang and its abundance increased due to subsequent stellar nucleosynthesis. During this process heavier elements are being produced alongside with helium which means that amount of it is strictly connected to amount of metals in interstellar medium. Determination of primordial helium abundance \(Y_p\) is of a special interest as it is an independent test to SBBN theory which predicts this value as a function of the baryon-to-photon ratio \(\eta\). Baryon-to-photon ratio in its own case is related to baryon density parameter \(\Omega_b\) by the expression \(10^{10} \eta = 273.9 \Omega_b h^2\), where \(h = H_0/100 \text{ km s}^{-1}\) and \(H_0\) is Hubble constant.

On the other hand \(Y_p\) depends sensitively on two physical constants: neutron lifetime \(\tau_n\) and number of neutrino species \(n_{\nu}\), so helium abundance is sensitive to non-standard early
Universe expansion rate and because of this it can provide strict constraints to non-standard physical processes taking place after the Big Bang. To detect such small deviations from SBBN theory and find appropriate constraints one have to determine $Y_p$ with accuracy better than one percent. There are different methods of $Y_p$ determination in different astrophysical objects, but the most traditional and precise way is to derive this quantity from observation of optical and near infrared (hereafter NIR) spectra of extragalactic regions which have undergone chemical evolution and thus have their composition close to the primordial one. Mainly these objects are blue compact dwarf (BCD) galaxies and HII regions with low metallicity (i.e., significantly below solar). To obtain required level of accuracy one must use high quality spectra and big sample of objects to maximally reduce statistical and instrumental uncertainties. Such sample of a total of 86 low-metallicity extragalactic HII regions in 77 galaxies had been gathered and introduced in [1] and is called HeBCD dataset.

While isotopic composition of these objects is similar to primordial one, actually it differs from SBBN composition and thus the primordial helium abundance is determined by fitting the helium abundance $Y$ versus metallicity extrapolating back to zero metallicity as described in [2]. First step to determine accurate values of $Y$ after getting high signal-noise ratio spectra is to correct this spectra for several known systematic effects. All of them are taken in account using special analytical fits for different physical processes, i.e. HeI emissivities or effects that makes observed line fluxes deviate from just recombination values like collisional excitation. Mainly there are six systematic effects that should be considered in derivation of $Y$: (1) HeI emissivities; (2) interstellar reddening; (3) the temperature structure of the HII region, i.e., the temperature difference between $T_e$(HeI) and $T_e$(OIII); (4) underlying stellar HeI absorption; (5) collisional excitation of hydrogen lines; (6) the ionization structure of the HII region. Detailed discussion of role of each of these effects is presented in [1].

Using HeBCD sample and new set of HeI line emissivities by Porter et al. [3, 4], Izotov et al.[5] and Aver et al. [6] subsequently derived following $Y_p$ values: $Y_p = 0.2551 \pm 0.0022$ by Izotov et al. and $Y_p = 0.2449 \pm 0.0040$ by Aver et al. These results are notably different from each other and do not coincide even within the margin of uncertainties. Considering both of the groups have been using same dataset and correcting it for similar systematic effects this fact is somewhat of a strange result. Aver’s et al. result is broadly consistent with $Y_p$ derived from baryon density parameter $\Omega_b$ based on CMB temperature fluctuation spectrum from Planck Observatory. The value of the primordial helium abundance derived that way is $Y_p = 0.2467 \pm 0.0006$ [7]. Both of this groups used in some way similar self-consistent methods to correct for systematic effects and derive $Y_p$ which will be described later on but results differ significantly. On the other hand Peimbert et al. [8] presented completely different and more deep study of several appropriate objects in order to derive $Y_p$ value. While two aforementioned groups used large sample of data of optical an NIR spectra to reduce statistical biases (28 and 15 objects for Izotov et al. and Aver et al. respectively), Peimbert et al. use only 5 objects and do not use NIR recombination lines, but they make a tailor-made model for each of this objects to maximally reduce systematic uncertainties. Peimbert et al. derive the primordial helium abundance as follows $Y_p = 0.2446 \pm 0.0029$ which is in good agreement with [6] result. Peimbert et al. compare their results with already named ones and discuss why they differ from each other, however they do not remark which of the groups is wrong and why.

In this paper we present comparison between two models from [5] and [6]. Because $Y_p$ value is linear extrapolation of $Y$ versus metallicity O/H and Izotov et al. $Y$ values of same objects are systematically higher compared to Aver et al. ones we try to find out the differences in models through derivation of $Y$. To do that we generate test synthetic spectra with known physical parameters and reproduce ones by two named methods to see if there are any difference. Paper is organized as follows. First, Sec. 2 introduces models that should be tested with particular emphasis on systematic effects accounting. Second, in Sec. 3 Monte-Carlo testing with synthetic
spectra is carried out to examine diagnostic power of two models. Finally, Sec. 4 offers a discussion of obtained results and possible ways of making models’ testing results converge to each other.

2. Models overview
As previously discussed the derived He mass fraction Y is somewhat higher in Izotov et al. analysis. Mass fraction Y is connected to density fraction y with formula:

\[ Y = \frac{4y(1 - B \cdot [O/H])}{1 + 4y} \]  

(2.1)

where \([O/H]\) is the oxygen to hydrogen mass fraction and \(B\) is special constant. Izotov et al. take \(B\) as a function from \([O/H]\) and thus \(B\) varies from 15 to 25, while Aver et al. simply take \(B = 20\). This shouldn’t make significant difference because \([O/H] \sim 10^{-5}\) and this term is small. Definition of \(y\) is significantly different in these two models. Izotov et al. [5] use following expression

\[ y = ICF(He) \times (y^+ + y^{2+}) \]  

(2.2)

where \(y^+ = He^+/H^+, \; y^{2+} = He^{2+}/H^+, \; ICF(He) = H^+/(He^+ + He^{2+})\) is the ionisation correction factor for He. This systematic effect comes from assumption that sizes of \(H^+\) and \(He^+\) zones in HII region can be of different sizes depending on hardness of the ionizing radiation. Aver et al. suggest amplitude and uncertainty in ICF to be much smaller than the other investigated factors, so they take \(ICF(He) = 1\) and \(y\) is simply \(y^+ + y^{2+}\) derived from Monte-Carlo analysis. \(ICF(He)\) derived in [5] are close to 1, and thus could change \(y\) maximum by 1-2\% which corresponds to changes only in third unit after the point and does not explain the significant differences in \(Y\). Now we came to the point where \(y^+\) should be determined and from this point methods suffer great diversification. Hereafter we will name model of Izotov et al. as ITG14 and model of Aver et al. as AOS15.

2.1. ITG14 model
5 strongest optical recombination lines and 1 NIR line are used to derive \(y^+\): \(\lambda 3889, \lambda 4471, \lambda 5576, \lambda 6678, \lambda 7065\) and \(\lambda 10830\). Derived \(y^+\) value in ITG14 model depend on following physical parameters: the fraction \(\Delta I(H\alpha)/I(H\alpha)\) of the \(I(H\alpha)\) emission-line flux due to collisional excitation, the electron density \(n_e(He^+)\), the electron temperature \(t_e(He^+)\), and the optical depth \(\tau\) of the \(HeI\) \(\lambda 3889\) emission line. Also input spectra firstly is corrected for reddening and underlying stellar absorption which involves \(a_{He}\) and \(a_H\) - the equivalent widths of the \(H\alpha\) and \(HeI\) underlying absorption lines and reddening parameter \(C(H\beta)\). To calculate \(y^+ \Delta I(H\alpha)/I(H\alpha), \; n_e(He^+), \; t_e(He^+)\) and \(\tau\) are simultaneously and randomly varied within specified ranges. For each value of \(\Delta I(H\alpha)/I(H\alpha)\; C(H\beta)\) and \(a_{He}\) are calculated using shifted Balmer decrement then \(HeI\) spectral lines are corrected for reddening and underlying absorption using expression

\[ \frac{I(\lambda)}{I(H\beta)} = \frac{F(\lambda)}{F(H\beta)} \frac{W(\lambda) + a_{He}(\lambda)}{W(H\beta)} \frac{W(H\beta)}{W(H\beta) + a_H(H\beta)} 10^{0.4(f(\lambda)C(H\beta))} \]  

(2.3)

where \(I(\lambda)\) is intrinsic line flux, \(F(\lambda)\) is observed line flux, \(W(\lambda)\) is the equivalent width of specified spectral line, \(f(\lambda)\) is reddening function. In ITG14 underlying absorption for \(He(4471)\) line is frozen and chosen to be 0.4 A. To calculate \(a_{He}\) for other lines this value is multiplied by coefficients listed in [5]. After these corrections have been carried out one should minimize following quantity to find best solution in multiparameter space with Monte-Carlo method:

\[ \chi^2 = \sum_{i=1}^{6} \frac{(y_i^+ - y_{wm})^2}{\sigma(y_i^+)^2}, \; \text{where} \; y_{wm} = \frac{\sum_{i=1}^{3} y_i^+ / \sigma(y_i^+)^2}{\sum_{i=1}^{3} 1 / \sigma(y_i^+)^2} \]  

(2.4)
Here $y_i^+$ is helium abundance derived from $i$ line flux and $\sigma(y_i^+)^2$ is its statistical error, $y_{\text{.emplace}}$ is the weighted mean of the $He^+$ abundance and it is the required quantity. It is important to note that all 6 lines are used to calculate $\chi^2$ functional while only lines of 4471, 5876 an 6678 Å are used in $y_{\text{emplace}}$ calculation. This is reasonable, because uncertainties of other lines are much greater compared to named ones, and this move results in more precise minimization [1]. Calculation of $y_i^+$ is done using following expression:

$$
y_i^+ = \frac{I(\lambda)}{I(H\beta)} \frac{\epsilon_{H\beta}(n_e, t_e)}{\epsilon_{\lambda}(n_e, t_e)} \frac{1}{f_\lambda(n_e, t_e, \tau) + C/R(\lambda)}
$$  \hspace{1cm} (2.5)

where $\epsilon_j$ is $j$-line emissivity from [4] which analytical fits are provided in [10], $f_\lambda(n_e, t_e, \tau)$ is optical depth function taken from [9] and $C/R$ is collisional to recombination ratio of $HeI$ emission lines with analytical fits from [10]. More detailed discussion of this model and accounting of systematic effects in it are presented in [1].

2.2. AOS15 model

Determination of $y^+$ in this model involves one more $HeI$ recombination line in addition to lines from ITG14 - $\lambda$4026, which is crucial for correct determination of $a_{He}$ [12], and 3 Balmer lines ($H\alpha, H\gamma, H\delta$), all of which are normalized to $H\beta$. Correction for all named systematic effects and counting physical parameters are made in one self-consistent minimization procedure to reduce possible statistical discrepancies arising during separate calculations. Other important point of this model compared to ITG14 one is that $y^+$ is not defined implicitly, as the average of six individual line abundances, instead it is one of input parameters, like temperature or density.

In AOS15 input parameters are used to calculate synthetic flux and compare it to the observed one, allowing for more standard $\chi^2$ definition:

$$
\chi^2 = \sum_\lambda \left( \frac{F(\lambda) - F(H\beta)}{\sigma(\lambda)} \right)^2
$$  \hspace{1cm} (2.6)

where the $He$ flux at each wavelength $\lambda$ relative to the flux in $H\beta$ is given by

$$
\frac{F(\lambda)}{F(H\beta)} = y^+ \frac{\epsilon_\lambda(n_e, t_e)}{\epsilon_{H\beta}(n_e, t_e)} f(n_e, t_e, \tau) \frac{W(H\beta) + a_{H}}{W(H\beta) + a_{He}(\lambda)} \frac{1 + C/R(\lambda)}{1 + C/R(H\beta)} 10^{-f(\lambda)C(H\beta)}
$$  \hspace{1cm} (2.7)

The ratio of $H$ fluxes is defined analogously

$$
\frac{F(\lambda)}{F(H\beta)} = \frac{\epsilon_\lambda(n_e, t_e)}{\epsilon_{H\beta}(n_e, t_e)} \frac{W(H\beta) + a_{H}}{W(H\beta) + a_{He}(\lambda)} \frac{1 + C/R(\lambda)}{1 + C/R(H\beta)} 10^{-f(\lambda)C(H\beta)}
$$  \hspace{1cm} (2.8)

Fits for collisional-to-recombination ratio for $He$ and $H$ lines are presented in [11]. It should be remarked that involving $C/R$ ratio for hydrogen lines explicitly introduces new physical parameter - neutral-to-ionized hydrogen density fraction, $\xi$. Emissivities are taken from [4] and [11]. Finally, in AOS15 model there are eight parameters: $y^+, n_e, t_e, \tau, a_{H}, a_{He}, C(H\beta)$ and $\xi$, which is two times more than in ITG14 one. As well as in ITG14, underlying absorption widths on different $\lambda$ are defined as functions of $a_{H}(H\beta)$ and $a_{He}(4471)$ with coefficients listed in [6]. Also weak temperature prior in form of $X_T = (T - T(OIII))^2 / (0.2 \cdot T(OIII))^2$ is added to (2.6). Minimization is done using Markov Chain Monte-Carlo analysis which allows to self-consistently determine named parameters with high precision.
3. Diagnostic power of presented models
Discrepancy in results between two independent groups could have risen because they were using different methods of $Y_p$ determination. From [5] and [6] it is easily seen, that for same sources Izotov et al. gives systematically higher values of $Y$, which apparently leads to higher $Y_p$. In order to understand which of methods returns better values of physical parameters we should test them on synthetical spectra with known parameters.

Syntactical spectra is composed of fluxes and errors for all named helium and hydrogen lines, plus their equivalent widths. Fluxes are generated based on known physical parameters with formulas from [6]. Errors are fixed to be 0.05 of flux value for hydrogen lines and 0.03 to 0.15 for helium lines. This is done to test models reactions to growing uncertainties, because 3% and less accuracy of line flux determinations are quite difficult to obtain. Equivalent widths are taken as a mean values of ones from real sources from HeBCD sample for each line. After generating this spectrum is given to two programs executing methods discussed previously. Results of their work are presented in tables 1 and 2. Uncertainties are taken on 95% confidence level.

Table 1. Results of processing synthetical spectra using AOS15.

| Parameters | Input | $\sigma_{He} = 0.03$ | $\sigma_{He} = 0.05$ | $\sigma_{He} = 0.1$ | $\sigma_{He} = 0.15$ |
|------------|-------|---------------------|---------------------|---------------------|---------------------|
| $y^+$      | 0.08  | 0.0794+0.0037-0.0025| 0.0810+0.0042-0.0013| 0.0801+0.0051-0.0056| 0.0798+0.0055-0.0055|
| $n_e$      | 100.0 | 97.8813+18.7766-18.1641 | 80.8866+19.7040-20.7288 | 99.2356+17.4183-18.4236 | 99.1919+15.2429-16.2660 |
| $t_e$      | 1.8   | 1.7930+0.1386-0.1386 | 1.8392+0.2064-0.2064 | 1.7026+0.3064-0.3064 | 1.6702+0.3577-0.3577 |
| $\tau$     | 0.2   | 0.3160+0.2302-0.2302 | 0.4873+0.3440-0.3440 | 0.6215+0.4529-0.4529 | 0.7596+0.5744-0.5744 |
| $a_H$      | 1.5   | 1.9045+1.1588-1.1588 | 2.1083+1.3436-1.3436 | 2.0797+1.6026-1.6026 | 1.9073+1.6574-1.6574 |
| $a_{He}$   | 0.4   | 0.4765+0.3036-0.3036 | 0.6399+0.4392-0.4392 | 0.6493+0.4969-0.4969 | 0.6827+0.5372-0.5372 |
| $C(H\beta)$| 0.25  | 0.2434+0.0271-0.0271 | 0.2288+0.0432-0.0432 | 0.2331+0.0530-0.0530 | 0.2417+0.0597-0.0597 |
| $\xi \times 10^4$ | 1.0 | 5.1295+3.3279-3.3279 | 5.0711+3.4515-3.4515 | 5.1517+3.5700-3.5700 | 5.0330+3.4002-3.4002 |
| $\chi^2$  | 0.0   | 0.1626              | 0.3542              | 0.5071              | 0.4738              |

Table 2. Results of processing synthetical spectra using ITG14.

| Parameters | Input | $\sigma_{He} = 0.03$ | $\sigma_{He} = 0.05$ | $\sigma_{He} = 0.1$ | $\sigma_{He} = 0.15$ |
|------------|-------|---------------------|---------------------|---------------------|---------------------|
| $y^+$      | 0.08  | 0.0776±0.0011        | 0.0764±0.0019        | 0.0795±0.0039        | 0.0826±0.0051        |
| $n_e$      | 100.0 | 417.71+0.090          | 373.089+0.2035       | 284.30+0.390         | 214.46+0.46          |
| $t_e$      | 1.8   | 1.851±0.009           | 1.827±0.014          | 1.81±0.017           | 1.805±0.018          |
| $\tau$     | 0.2   | 1.206±0.065           | 1.438±0.121          | 1.572±0.15           | 1.615±0.176          |
| $a_H$      | 1.5   | 0.978±0.19            | 0.988±0.19           | 1.001±0.208          | 1.009±0.205          |
| $C(H\beta)$| 0.25  | 0.281                | 0.279                | 0.279                | 0.279                |
| $\chi^2$  | -     | 87.172               | 33.99                | 8.646                | 3.858                |

As seen from results both methods give reasonably close estimation for synthetical spectrum parameters which grows with growing errors of flux measurements. Tests show that AOS15 model returns more precise values but partly this is due to way the synthetic spectrum is generated. But method is statistically stable and robust. On the other hand ITG14 model...
returns values that differ a lot from given ones. As seen from table 2 ITG14 model underestimates $y^+$ strongly which is quite weird, especially knowing the fact that when given real source spectra ITG14 model returns statistically higher $y^+$ values, compared to AOS15. Because in synthetic spectra all lines have same uncertainty, all of them give comparable contribution to $\chi^2$ sum and to weighed mean value of $y^+$. When real-life sources are taken uncertainties differ a lot from line to line, which results in more correct calculations.

4. Conclusion

Primary aim of this article was to test diagnostic power of two different methods compared to each other and try to understand pros and cons of them in order to develop more general method of $Y_p$ determination later, which could include strong sides of both models.

AOS15 model shows great efficiency of parameter determination for low-errored flux, and discrepancy grows with increasing errors, which means that this model can return highly underestimated result for inexact spectra. This model achieves its best accuracy for sources with less than 3% error in fluxes. Main problem for this model is that in HeBCD relative errors of line fluxes are bigger than 5%, whooping shooting up to almost 40% for certain lines which as been shown leads to underestimation of helium content in a source.

On the other hand ITG14 models shows greater stability for increasing flux errors, but it almost always gives biased evaluations for all physical parameters which could lead to overestimated constraints of helium determination. Probably it is caused by non-self-consistent procedure or by fixed underlying absorption for helium. It is seen from table 2, that most biased parameter is optical depth, but for ITG14 way of calculating $y^+$ only 4471, 5876 and 6678 Ålines are taken in final analyses, where optical depth function from [9] is very close to 1. It is seen from table 2 that value of $y^+$ grows with uncertainty increment. It could result in overestimation of helium concentration for highly uncertain spectra, but this statement needs more investigation. Problems of ITG14 method could be solved by inclusion of some extra lines like 4026 Å or additional parameters in $\chi^2$ function.

Though tests of diagnostic power on synthetical spectrum showed satisfying results for both procedures of evaluating $y^+$ in a source, we must take in account other differences in two named methods. For example, one of the main differences that could have introduced additional uncertainty is way the HeBCD sample is filtered for qualifying objets. Taking in account all this statements we conclude that diagnostic power is fair but could be improved and we can’t strictly state one of to models to be preferable. Other details should be considered and additional research should be done on this problem.

Acknowledgments

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

References

[1] Izotov, Y. I. and Thuan, T. X. and Stasińska, G. 2007, ApJ 662 15-38
[2] Pagel, B. E. J. and Simonson, E. A. and Terlevich, R. J. and Edmunds, M. G. 1992, MNRAS 255 325-345
[3] Porter, R. L. and Ferland, G. J. and Storey, P. J. and Detisch, M. J. 2012, MNRAS B 425 L28-L31
[4] Porter, R. L. and Ferland, G. J. and Storey, P. J. and Detisch, M. J. 2013, MNRAS B 433 L89-L90
[5] Izotov, Y. I. and Thuan, T. X. and Guseva, N. G. 2014, MNRAS 445 778-793
[6] Aver, E. and Olive, K. A. and Skillman, E. D. 2015, JCAP 7 11
[7] Planck Collaboration and Ade, P. A. R. and Aghanim, N. and Arnaud, M. and Ashdown, M. and Aumont, J. and Baccigalupi, C. et al. 2016, AAp 594 A13
[8] Peimbert, A. and Peimbert, M. and Luridiana, V. 2016, RMAA 52 419-424
[9] Benjamin, R. A. and Skillman, E. D. and Smits, D. P. 2002, ApJ 569 288-294
[10] Izotov, Y. I. and Stasińska, G. and Guseva, N. G. 2013, AAp 558 A57
[11] Aver, E. and Olive, K. A. and Skillman, E. D. 2010, JCAP 5 003
[12] Aver, E. and Olive, K. A. and Skillman, E. D. 2011, JCAP 16 003