HII-CHI-mistry: A collection of phython scripts for the analysis of emission lines in ionized gaseous nebulae

Enrique Pérez-Montero

Instituto de Astrofísica de Andalucía - CSIC. Apdo. 3004, 15080, Granada, Spain

1. General description

1.1. *What is* HII-CHI-mistry *and where can it be found?*

Ionized gaseous nebulae are ubiquitous objects in the Universe that provide information about the galaxies where they are located by means of their bright emission lines. Among the different properties that we can derive from them are the relative chemical abundances of the observed ions, the excitation of the gas or the effective hardening of the ionizing source. HII-CHI-mistry (hereinafter, HCm) is a collection of python scripts that help to analyze the observational information from several bright emission lines observed in the ultraviolet, optical, or infrared ranges of the spectrum, comparing them with the predictions from large grids of photoionization models and providing estimates with their errors for some derived properties, adapting the accuracy of the solutions to the used input lines and their uncertainties.

HCm for the calculation of chemical abundances presents three main advantages in relation to other model-based solutions, including: 1) the results are totally consistent with the so-called direct method, based on the previous determination of the electron temperature. This is true even when the flux of an auroral line, such as [[O III]] λ 4363 Å is not used (see Pérez-Montero, 2014). 2) HCm provides consistent solutions independently of the set of input emission lines and their errors. This is especially useful when we want to compare results for different sets of objects observed with different spectral ranges or redshift. 3) As HCm provides independent solutions for N/O (in the optical and the infrared) or C/O (in the ultraviolet), , it allows the used from N or C for the derivation of O/H without any previous assumption about the relation between O/H and N/O or C/O.

In this tutorial I explain the features of the different versions of the code and its instructions of use, along with its limitations and future expected improvements. There are four versions of HCm depending on the spectral range or its use for the determination of chemical abundances or the calculation of the hardening of the ionizing source. These are

- HII-CHI-mistry. The original HCm package described in Pérez-Montero, 2014 for the analysis of gaseous nebulae ionized by massive young star clusters and in Pérez-Montero et al., 2019a for Narrow Line Regions (NLR) ionized by active galactic nuclei (AGN), used to derive the total oxygen abundance (12+log(O/H), hereafter O/H), the nitrogen-to-oxygen chemical abundance ratio (log(N/O), hereafter N/O), and the ionization parameter (log U) using optical emission lines from [[O II]] λ 3727 Å up to [[S II]] λ 6717+6731 Å.
- HII-CHI-mistry-UV. Described in Pérez-Montero and Amorín, 2017 for star-forming objects, and in Pérez-Montero et al., 2023a for the NLR of AGN. It calculates O/H, the carbon-to-oxygen chemical abundance ratio (log(C/O), hereafter C/O), and log U, using ultraviolet emission lines from Lyα λ 1216 Å up to C III] λ 1909 Å.
- HII-CHI-mistry-IR. Described in Fernández-Ontiveros et al., 2021 for star-forming

objects, and in Pérez-Díaz et al., 2022 for the NLR of AGNs. It calculates O/H, N/O, and log U using infrared emission lines from Brα λ 4.05 μm up to [[N II]] λ 205 μm.

• HII-CHI-MISTRY-TEFF. It calculates the equivalent effective temperature (T_*) and log U using optical and UV emission lines and O/H, as described in Pérez-Montero et al., 2019b. It can also be used to derive the fraction of absorbed ionizing photons for objects with He II λ 4686 Å as described in Pérez-Montero et al., 2020.

• HCM-TEFF-IR. It calculates T_* and U based on near- and mid-infrared emission-lines. It is described in Pérez-Montero et al (2023c).

All versions of HCM have been written in `python` and should work for both versions 2 and 3. All can be downloaded from the webpage of HCM at http://www.iaa.es/ epm/HII-CHI-mistry.html. In this document I describe features for the public versions of HCM 5.3 and equivalents for the other versions, which require the `python` library `numpy`. Each package contains the corresponding script (named with `.py` extension), an `ascii` file with instructions, an input text file example, and different libraries containing the information from the models, and template files with information about constrained sub-grids used by the code in absence of certain emission-lines. All files are included in a compressed `tgz`[o] file. These can be uncompressed using from a terminal prompt the command

```
> tar xvfz HCm_v5.3.tar.gz
```

1.2. *How does the code work?*

All versions of HCM performs a bayesian-like approach to derive the chemical abundances, ionization parameter or T_*. In brief, for a given property X, the final result is

$$X_f = \frac{\sum_i X_i/\chi_i}{\sum_i 1/\chi_i} \tag{1.1}$$

where X_f is the result, X_i are the input values in each one of the models of the grid, and χ are the weights assigned to each one of the models, calculated as the quadratic difference between the observed and the predicted values for some specific emission-line ratios

$$\chi_i^2 = \sum_j \frac{(O_j - T_{ji})^2}{O_j} \tag{1.2}$$

being O_j and T_{ji} the observed and model-based values, respectively, for the considered emission-line dependent ratios. These are described in each one of the papers of the different versions of HCM as a function of the input emission lines.

The errors assigned to each result is calculated as the quadratic sum between the dispersion of all results obtained following a Monte-Carlo iteration through the nominal values perturbed with the nominal input observational errors, and the mean of all intrinsic uncertainties assigned to the bayesian process, calculated as:

$$(\Delta X)^2 = \frac{\sum_i X_f - X_i)^2/\chi_i}{\sum_i 1/\chi_i} \tag{1.3}$$

For those versions of HCM aimed at the calculation of chemical abundances, the first iteration is used to provide an estimation for the abundance of a secondary ion (N for

HCm and HCm-Ir, or C for HCm-UV),relative to oxygen, as these are based on emission-line ratios very low sensitive to log U. This also has the advantage that the grid can be constrained in a second iteration, once N/O or C/O are fixed, to calculate both O/O and log U using N or C lines without any a-priori assumption for the relation between O and the respective secondary element.

In the case of HCm-Teff, both for optical and IR, this previous iteration is not performed, but the grid of models is interpolated to fix the value of O/H to the input value in order to minimize the dependence of both $T :_*$ and U on metallicity.

2. HCm in the optical

2.1. *Running the program*

To run the HCm code in version 5.3, we must typed in the terminal prompt:

```
> python HCm_v5.3.py
```

A text describing the code will appear on the screen and it will ask for the input file containing the observational information. Alternatively, this name can be typed when the code is invoked, along with the desired number of iterations for the Monte-Carlo simulation:

```
> python HCm_v5.3.py input.txt 100
```

When the number of iterations is not specified, this is put by default to 25.

2.2. *The input file*

The input file must be an `ascii` file, with as many rows as the number of objects or pointing for which we want to perform the calculation. Each introduced column represents the identification for each row and the reddening corrected relative to Hβ fluxes with their corresponding errors. In previous versions of HCm, all columns should be introduced, but from version 5.0 a first row with the labels of the introduced columns is just required and the order is not essential. The file also admits other columns not identified by the code. The columns assigned to the observational relative errors are neither mandatory. The labels for the emission lines in this version are::

- `ID`. To identify each row with a name..
- `OII_3727` and `eOII_3727`, for [[O II]] λ 3727 Å and its error. This is assumed to be the addition of λ 3726 Å and λ 3729 Å in resolved spectra.
- `NeIII_3868` and `eNeIII_3868`, for [[Ne III]] λ 3868 Å and its error.
- `OIII_4363` and `eOIII_4363`, for [[O III]] λ 4363 Å and its error.
- `OIII_4959` and `eOIII_4959`, for [[O III]] λ 4959 Å and its error.
- `OIII_5007` and `eOIII_5007`, for [[O III]] λ 5007 Å and its error. When only one of the two strong [[O III]] lines is given, the code assumes the theoretical ratio between them into the account.
- `NII_5755` and `eNII_5755`, for [[N II]] λ 5755 Å and its error.
- `SIII_6312` and `eSIII_6312`, for [[S III]] λ 6312 Å and its error.
- `NII_6584` and `eNII_6584`, for [[N II]] λ 6584 Å and its error.
- `SII_6716` and `eSII_6716`, for [[S II]] λ 6716 Å and its error.
- `SII_6731` and `eSII_6731`, for [[S II]] λ 6731 Å and its error. Notice that the code will only use the [SII] lines if both of them of their addition is given.
- Alternatively, it can be given the addition of these two last with `SII_6725` and `SII_6725`.

- `OI_7325` and `eOII_7325`, for [[O II]] λ 73109+7330 Å and its error.
- `SIII_9069` and `eSIII_9069`, for [[S III]] λ 90699 Å and its error.
- `SIII_9532` and `eSIII_9532`, for [[O III]] λ 9532 Å and its error. When only one of the two strong [[S III]] lines is given, the code assumes the theoretical ratio between them into the account.

2.3. *Selecting the grid of models*

If the input file is correct, the code will ask for the grid of models to perform the calculation:

```
(1) POPSTAR with Chabrier IMF, age = 1 Myr
(2) BPASS v.2.1 a_IMF = 1.35, Mup = 300, age = 1Myr with binaries
(3) AGN, double component, a(UV) = -1.0

Other SED
) (4) Different library
```

`Choose SED of the models:`
All grids have been calculated using the code CLOUDY v.17 (Ferland et al., 2017) assuming a central ionizing source and a plane-parallel geometry. The grid 1 is described in Pérez-Montero, 2014 and is calculated using POPSTAR (Mollá et al., 2009) cluster model atmospheres with instantaneous burst at an age of 1Myr, assuming a Chabrier, 2003 initial mass function (IMF) and a constant electron density of 100 cm^{-3}. The grid 2 is described in Pérez-Montero et al., 2021 and it was calculated with cluster model atmospheres from BPASS v.2.1 (Eldridge et al., 2017), assuming an instantaneous burst at an age of 1 Myr and an IMF with slope x = -1.35 and with an upper mass limit of 300 M$_\odot$ and with binaries. The gas in the models has an electron density of 100 cm^{-3}. This grid is advisable for the case of Extreme Emission Line Galaxies (EELGs).

The option 3 can be used to derive chemical abundances in the NLR of AGN and include several grids which are described in Pérez-Montero et al., 2019a. These assume a doubled peak power law spectral energy distribution with a parameter α_{UV} = -1.0 . The assumed electron density is 500 cm^{-3}. The code will ask for the value considered in these models for α_{OX} (-0.8 or -1.2) and the criterion used to stop the models (at an outer radius when the fraction of free electrons is 98% or 2%).

All default grids cover an input O/H value in the range [6.9,9.1] in bins of 0.1 dex and N/O in the range [-2.0,0.0] in bins of 0.125 dex. Regarding log U is covered with a resolution of 0.25 dex in the range [-4.0,-1.5] for those grids for star-forming regions (i.e. 1 and 2= and in the range [-2.5,-0.5] for the grids for AGN. The user can also use his/her model libraries, including them as a text file in the correct format in the folder `Libraries_opt`. In case a different library is introduced by the user (4), the code will automatically check if the file shows the correct format. In case some of the information is missing the code will warn the user about the missing columns.

Once the grid of models has been selected, the code will ask for the use of interpolation between the models of the grid:
`Choose models` [0] `No interpolated` [1] `Interpolated:`
In the interpolated mode, the code makes a linear interpolation of all variables multiplying the resolution by a factor 10. This precludes the results clustering around certain knots of the grid. On the contrary, this mode slows the calculation time.

As a final step, the program will ask for the constraint laws that will be used to limit the grids. This is necessary when a limited set of emission lines is given and the models have to take assumptions on the relations between O/H and N/O (i.e. when N/O cannot

be derived) or O/H and U (i.e. when no auroral line is given, so excitation is used to estimate metallicity). For AGNs as for the possible input lines there is a degeneracy of the O2O3 emission-line ratio as discussed in Pérez-Montero et al., 2019a, to derive $\log U$, the code will also ask for the preferred $\log U$ range selected for the calculations between high-ionization ($\log U \geq$ -2.5) or low-ionization ($\log U \leq$ -2.5). .

```
Select a file with the constraint laws to be used to limit the grid of models
when the measurement of a quantity is impossible without any relation.

Default constraints

(1) Constraints for Star-Forming Galaxies
(2) Constraints for Extreme Emission Line Galaxies
(3) Constraints for AGNs (no restriction in the ionization parameter)
(4) Constraints for high ionization AGNs (log(U) > -2.5)'
(5) Constraints for law ionization AGNs (log(U) < -2.5)

(") Other constraints

('———') (6) Different constraint file'
```

The user can also use his/her own constrain assumption in a text file with the correct format in the folder `Constrains`. In case that)the option 6 is chosen ,the code will check the file for the correct format. In case some of the information is missing , the code will warn the user about the missing columns.

After this process, the program will summarize the different grids that are going to be used and the number of models for each grid.

2.4. *Results*

The code will show on the screen the results for the calculations for each one of the rows, showing as well the ratio of completeness of the task. At the end, it will create another `ascii` file with all the results named in the same way as the input file, but adding `_hcm-output.dat`. The first column lists the identification of each row. If this has not been specified in the input file, a number is assigned. The next columns of this file contain the emission-line fluxes used as input, with their corresponding errors.

The next column is an index that indicates if the whole grid of models has been used or, in contrast, a constrained grid has been used instead. This depends mainly in each row on the introduced emission-lines, as a limited observational set implies additional assumptions to calculate ionic abundances but, as described above, these constrains can be chosen as a function of the models or instead the user can use his/her own. An index 1 indicates that the whole grid of models has been used. This is done when a estimation of the electron temperature, similarly as in the direct method (see for instance Pérez-Montero, 2017) can be made because both auroral and nebular emission lines are present (i.e. [[O III]] 4363 Å with 4959,5007 Å, [[N II]] 5755 Å with 6584 Å, [[O II]] 7325 Å with 3727 Å, or [[S III]] 6312 Å with 9069,9532 Å). If [[O III]] λ 4363 Å is not given, what it is a common situation in fain or metal-rich nebulae (Pérez-Montero and Díaz, 2005), additional assumptions should be made.

The index 2 then represents the case when electron temperature cannot be estimated and the code assumes an empirical law between O/H and $\log U$ (see Figure 1). This relation assumes that metal-poor objects present in average higher excitation and, on the contrary, metal-rich objects have lower excitation. This is the same assumption behind the

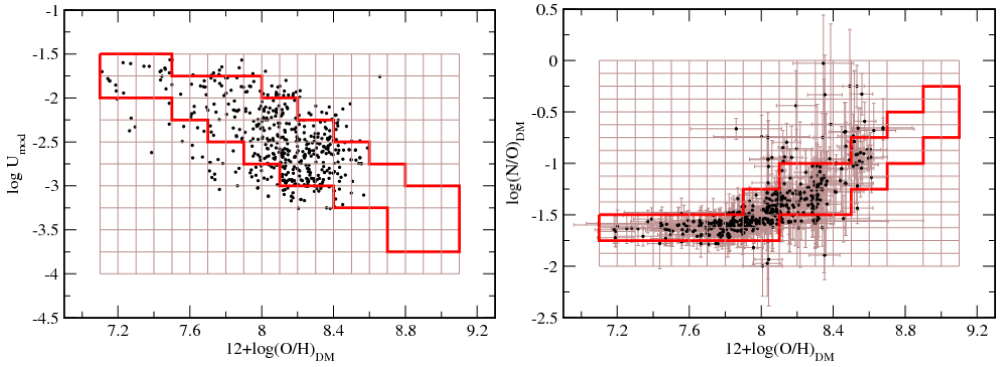

FIGURE 1. At left, relation between total oxygen abundance and ionization parameter for the sample studied in Pérez-Montero, 2014. The solid red line encompasses the most probable combination of parameters occupied by the objects. At right, empirical relation between 12+log(O/H) and log(N/O) with the region occupied by the grid of models in case of no available observational information to constrain N/O.

use of many high- to low-excitation emission line flux ratios (e.g. O3O2, O3N2) to derive chemical abundances. Finally, index 3 denotes the use of a grid constrained assuming an empirical relation between O/H and N/O as shown in Figure 1. This is necessary in the case that N/O cannot be calculated independently in a first iteration using emission-line ratios such as N2O2, O3N2 or N2S2. This assumption is behind all strong-line calibrations based on [[N II]] and implies a constant N/O value for low metallicities due to a mostly primary production of N, and a increasing N/O with O/H when a certain production of secondary N is assumed. This assumption can lead to non-negligible deviations from the real O/H if N/O does not lie in the expected regime, as discussed in Pérez-Montero and Contini, 2009. The final six columns of the output file give the results for O/H, N/O, and log U with their corresponding errors. If no solution is found, O/H and log U are denoted by 0, and N/O is denoted by -10

The results and consistency for the abundances derived from HCM both forstar-forming objects whose abundances were derived from the direct method (Pérez-Montero, 2014) and for NLR in AGN as dervied using detailed photoionization models (Pérez-Montero et al., 2019a) are well discussed. Nevertheless, in Table 1 I provide a list of the mean offsets and the standard deviation of the residuals of the resulting O/H, N/O, and log U as compared with the input values from the models as a function of the used emission lines. These values illustrate how the code can recover the abundances from the values using only the emission-lines as input, but cannot be taken as true uncertainties.

3. HCM in the ultraviolet

The HII-CHI-MISTRY-UV code (hereafter HCM-UV) is similar to the version in the optical but using a different set of emission lines in the ultraviolet regime and estimating in a first iteration C/O instead N/O. It also admits two emission lines in the optical in order to provide an estimate of the abundance based on an emission-line ratio sensitive to the electron temperature (e.g. [[O III]]λ 5007/λ 1665), although this can imply a larger inaccuracy due to reddening uncertainties. This code is described in Pérez-Montero and Amorín, 2017 for star-forming objects, and in Pérez-Montero et al., 2023a for the NLR in AGN.

TABLE 1. Mean offsets and standard deviation of the residuals of the resulting properties derived by HCM when the model from POPSTAR emission lines are used as input as a function of the used emission-line ratios. [[O III]]$_a$ stands for the [[O III]] auroral line at λ 4363 Å and [[O III]]$_n$ for the nebular lines at λ 4959,5007 Å.

Used lines	Grid	O/H		N/O		log U	
		Mean Δ	σ res.	Mean Δ	σ res.	Mean Δ	σ res.
All lines	1	+0.03	0.09	-0.07	0.08	+0.06	0.15
[[O III]]$_a$,[[O III]]$_n$,[[N II]],[[S II]]	1	+0.02	0.09	-0.01	0.07	+0.06	0.15
[[O II]],[[O III]]$_n$,[[N II]],[[S II]]	2	-0.04	0.08	+0.00	0.03	+0.05	0.10
[[O III]]$_n$,[[N II]],[[S II]]	2	-0.03	0.07	+0.00	0.04	+0.05	0.10
[[O II]],[[O III]]$_n$,[[N II]]	2	-0.04	0.24	+0.00	0.23	+0.03	0.09
[[N II]],[[S II]]	2	+0.00	0.19	+0.00	0.04	+0.02	0.20
[[O III]]$_n$,[[N II]]	3	-0.04	0.08	–	–	+0.07	0.13
[[N II]]	3	+0.01	0.15	–	–	+0.03	0.21
[[O II]],[[O III]]$_n$	3	+0.00	0.01	–	–	+0.00	0.02
[[O II]],[[Ne III]]	3	-0.01	0.02	–	–	-0.01	0.03

3.1. *Running the program and preparing the Input file*

Similarly to the version in the optical, it can be executed from the terminal prompt with python and allows either just calling the script or to precise the name of the input file and the number of iterations for the Monte Carlo simulations:

```
> python HCm-UV_v5.0.py input.txt 100
```

The input file is an `ascii` file whose first row must the labels of the used columns, including:

- `ID`. Identification name for each row.
- `Lya_1216` and `eLya_1216`, for Lyα λ 1216 Å and its error.
- `NV_1239` and `eNV_1239` for N V] λ 1239 Å and its error.
- `CIV_1549` and `eCIV_1549` for C IV] λ 1549 Å and its error.
- `HeII_1640` and `eHeII_1640` for He IIλ 1640 Å and its error.
- `OIII_1665` and `eOIII_1665` for [O III]] λ 1665 Å and its error. This includes all lines of the [O III]] multiplet lines between 1660 Å and 1666 Å.
- `CIII_1909` and `CIII_1909` for C III] 1909 Å and its error. Again, this includes all the lines of the C III] multiplet.
- `Hb_4861` and `eHb_4861` for Hβ λ 4861 Å**AND** its error.
- `OIII_5007` and `eOIII_5007` for [[O III]] λ 5007 Å and its error.

The following rows must be the names and extinction corrected fluxes in arbitrary units. A value zero can be used for missing values and not all columns are mandatory to get an estimation for O/H, C/O or U.

If the input file is correct the code will ask for the grid of models and the use of interpolation. . For version 5.0, models from POPSTAR, BPASS, and double composite AGN are available. In the case of AGN, the code will ask for the assumed α_{OX} (-0.8 or -1.2), the stopping criterion assumed for the models (2% or 98% of free electrons), or the assumption of presence or absence of dust. According to Pérez-Montero et al (2023) models are only able to reproduce C3C4 values $<$ -0.3 when dust is not considered.

In the next step, the code will also ask for the interpolation for the models in order to enhance the resolution, and the choice of templates considering assumptions on the

TABLE 2. Mean offsets and standard deviation of the residuals of the resulting properties derived by HCM-UV when the model from POPSTAR emission lines are used as input as a function of the used emission-line ratios.

Used lines	Grid	O/H		C/O		log U	
		Mean Δ	σ res.	Mean Δ	σ res.	Mean Δ	σ res.
All lines	1	+0.02	0.28	-0.05	0.08	+0.04	0.25
Lyα, C IV],He II,[O III]],C III]	2	+0.02	0.27	-0.06	0.19	+0.05	0.10
C IV],He II,[O III]],C III]	2	+0.10	0.30	-0.14	0.13	-0.11	0.20
Lyα,C IV],He II,C III]	3	+0.00	0.02	–	–	+0.00	0.01
C IV],He II,C III]	3	-0.01	0.03	–	–	+0.00	0.02

relation between O/H, log U, and C/O in case that some of the input emission lines are missing. These templates can also be edited by the user in the appropriate folder.

3.2. *Output file and analysis of the results*

Once the grid of models is selected, the code will show on the screen the results for each one of the input rows, along with the ratio of completeness. At the end, it will create and output file in `ascii` format with the results, including all information chosen for the calculation. The first column is the corresponding ID and the next will be the input emission lines along with their errors. The next column is an index that informs about if the complete grid has been used (index 1), only when both [[O III]] λ 1665 and 5007 Å are given) to provide an estimate of the electron temperature. Index 2 denotes a grid with an implicit relation between O/H and log U as shown in Figure 1. Finally, index 3 denotes a grid constrained assuming a relation between C/O and O/H, similar to that sown in Figure 1 for N/O and O/H, considering a fixed C/N to the solar ratio. This is used when C/O cannot be calculated by means of the emission line ratio C3O3, depending on C III] 1909 Å and [O III]] λ 1665 Å but an estimate for both O/H and log U can be given using C lines. In any case, as commented above, this relation can be changed by the user in the folder `Libraries_uv`. The code, for instance, also provides the relation derived for EELGs given by Pérez-Montero et al., 2021. The six final columns give the results for O/H, C/O, and log U with their corresponding errors. As in the case for the optical, if no solution can be found for both O/H and log U these are denoted as 0 in the file, and -10 for C/O

In Table 2 I list the mean offsets and the standard deviation of the residuals when we use as input for the code the same predictions from the models of the grid, when POPSTAR models are used as a function of the different combination of emission lines that lead to a solution. A similar table for the results for AGN can be found in Pérez-Montero et al (2023).

4. HCM in the infrared

The program HII-CHI-MISTRY-IR (hereafter HCM-IR) calculates total oxygen abundance, nitrogen-to-oxygen abundance ratio, and the ionization parameter from a set of observed emission-lines in the mid infrared. From version 3.1 the program also supplies a solution for the sulphur abundances. It is described in (Fernández-Ontiveros et al., 2021. From version 3.0 the program also admits the calculation for the NLR in AGNs and it is described in Pérez-Díaz et al., 2022.

4.1. *Running the program and preparing the input file*

Similarly to the other versions of the program, HCm-IR is executed from a terminal prompt using python, and the input text file can be invoked in the same sentence, along with the number of iterations for the Monte Carlo simulations:

```
> python HCm-IR_v3.1.py input.txt 50
```

The input file must be typed in `ascii` format. The first row specifies the identification of each row and the input emission lines, with the following accepted labels:

- `ID`. For the identification name for each row.
- `HI_4m` and `eHI_4m` for Paα at λ 4.07 μm and its error.
- `ArIII_7m` and `ArIII_7m` for [[Ar III]] at λ 6.98 μm and its error.
- `HI_7m` and `HI_7m` for Brα at λ 7.46 μm and its error.
- `ArIV_8m` and `ArIV_8m` for [[Ar IV]] at λ 7.90 μm and its error.
- `ArIII_9m` and `ArIII_9m` for [[Ar III]] at λ 8.99 μm and its error.
- `SIV_10m` and `eSIV_10m` for [[S IV]] at λ at 10.5 μm and its error
- `HI_12m` and `HI_12m` for Huα at λ 12.46 μm and its error.
- `NeII_12m` and `eNeII_12m` for [[Ne II]] λ 12.8 μm and its error.
- `ArV_13m` and `ArV_13m` for [[Ar V]] at λ 13.1 μm and its error.
- `NeV_14m` and `eNeV_14m` for [NeV] λ 14.9 μm and its error.
- `NeIII_15m` and `eNeIII_15m` for [[Ne III]] λ 15.5 μm and its error.
- `SIII_18m` and `eSIII_18m` for [[S III]] λ 18.8 μm and its error.
- `NeV_24m` and `eNeV_24m` for [NeV] λ 24.3 μm and its error.
- `OIV_25m` and `eOIV_25m` for [OIV] λ 25.9 μm and its error.
- `SIII_33m` and `eSIII_33m` for [[S III]] λ at 33.7 μm and its error.
- `OIII_52m` and `OIII_52m` for [[O III]] λ 52 μm and its error.
- `NII_57m` and `eNII_57m` for [[N II]] λ 57 μm and its error
- `OIII_88m` and `eOIII_88m` for [[O III]] λ 88 μm and its error
- `NII_122m` and `eNII_122m` for [[N II]] λ 122 μm and its error.
- `NII_205m` and `eNII_205m` for [[N II]] λ 205 μm and its error.

Then the program will ask for the chosen grid of models. For version 3.0, models from POPSTAR, BPASS, and double composite AGN are available. In the case of AGN, the code will ask for the assumed α_{OX} (-0.8 or -1.2), and the stopping criterion assumed for the models (2% or 98% of free electrons). In addition, as in other versions, it is possible to incorporate other grids defined by the user, conveniently specified and stored in the `Libraries_ir` folder. The code will also ask for the possibility of model interpolation to increase by a factor 10 the resolution of the grid, but this can slow the time of calculation. Finally, the code will also ask for the constrains to consider when no all emission-lines can be used. Again, the code can provide templates calculated for star-forming galaxies by Pérez-Montero, 2014, by EELGs by Pérez-Montero et al., 2021, or assuming low- or high-excitation AGNs. In any case, again, other constrain files can be incorporated by the user.

4.2. *Output file and analysis of the results*

As in previous versions of HCm, once the grid of models has been selected, the code will show in the screen results for O/H, N/O, and log U along with the ratio of completeness of the task. In addition, from version 3.1 sulphur abundances are also calculated if S lines are provided. It will also create another `ascii` file whose first column is the identification of each row followed by the emission lines with their errors given as inputs. The next column indicates if the complete grid is used (index 1). However, as no auroral lines are available in this spectral range, this is not used in any case. The index 2 corresponds to

TABLE 3. Mean offsets and standard deviation of the residuals of the resulting properties derived by HCM-IR when the models from POPSTAR emission lines are used as input as a function of the used emission-line ratios.

Used lines	Grid	O/H		N/O		log U	
		Mean Δ	σ res.	Mean Δ	σ res.	Mean Δ	σ res.
All lines	2	+0.04	0.09	-0.00	0.01	+0.02	0.07
[[Ne II]],[[Ne III]],[[S II]],[[S IV]],[[O III]],[[N II]],[[N III]]	2	+-0.03	0.13	+0.00	0.01	+0.02	0.07
[[O III]],[[N II]],[[N III]]	2	-0.01	0.18	+0.00	0.01	+0.02	0.15
HI,[[S III]],[[S IV]],[[Ne II]],[[Ne III]]	3	+0.04	0.04	–	–	-0.01	0.01
[[S III]],[[S IV]],[[Ne II]],[[Ne III]]	3	+0.04	0.04	–	–	+0.00	0.01
[[S III]],[[S IV]]	3	+0.02	0.02	–	–	+0.00	0.01
[[Ne II]],[[Ne III]]	3	+0.03	0.09	–	–	+0.00	0.06
[[N II]],[[N III]]	3	+0.00	0.11	–	–	+0.00	0.06

the grid constrained following an empirical relation between O/H and log U, as shown in Figure 1. Finally, index 3 is used when no previous estimation of N/O, using the N3O3 parameter can be done, and an empirical relation between O/H and N/O is assumed as shown in right panel of Figure 1. Alternatively, these relations can be changed to EELGs or others defined by the user, as commented above.

In Table 3 I list mean offsets and standard deviations of the residuals for the resulting O/H, N/O, and log U derived from the code from different sets of input emission lines, as compared with the values used as input in each grid of the model.

5. HCM for the calculation of T_*

The HII-CHI-MISTRY-TEFF)hereafter HCM-TEFF is different to the previously described versions of HCM, as its aim is not the derivation of chemical abundances, but the calculation of the equivalent effective temperature of the ionizing source (T_* or T_{eff}) or the fraction of absorbed ionizing photons assuming a matter-bounded geometry (f_{abs}). This code makes use of the relation between the so-called softness parameter (Vilchez and Pagel, 1988, Pérez-Montero and Vílchez, 2009) with the hardness of the ionizing incident radiation, and the program is well described in Pérez-Montero et al., 2019b and Pérez-Montero et al., 2020. In addition, a discussion on the effect of the diffuse ionized gas (DIG) on the results when using certain low-excitation lines, such as [[S II]], is also discussed in Pérez-Montero et al., 2023b. An independent version for the Ir, only valid so far for the derivation of T_* and U is also available and it is described in Pérez-Montero et al (2023c).

5.1. *Running the program and preparing the input file*

Running the program is equivalent to the previous versions of HCM, through the terminal prompt:

```
> HCm-Teff_v5.2.py input.txt 50
```

The input file must be written in `ascii` format. The first row should correspond to the object ID and the labels of the used emission lines and, if availab le, the oxygen abundance, which can reduce the uncertainty.. The code admits the following labels:

- ID. identification name for each row.

• `12logOH` and `e12logOH` for the total oxygen abundance 12+log(O/H) and its error. Maybe this can be calculated using HCm in a previous iteration.
• `CIV_1549` and `eCIV_1549` for C IV] λ 1549 Å and its error.
• `CIII_1909` and `eCIII_1909` for C III] λ 1909 Å and its error. This should include all lines taking part of the multiplet around this wavelength.
• `OII_3727` and `eOII_3727` for [[O II]] λ 3727 Å and its error. As in the case of HCm this represents the addition of the two lines of the double of [[O II]] if there is good spectral resolution.
• `OIII_4959` and `eOIII_4959` for [[O III]] λ 4959 Å and its error.
• `OIII_5007` and `OIII_5007` for [[O III]] λ 5007 Å and its error. As in the case of HCm if only one of the two lines of the [[O III]] doublet 4959,5007 is introduced , the code assumes its addition taking into account its theoretical relation.
• `HeI_4471` and `HeI_4471` for He I λ 4471 Å and its error.
• `HeI_5876` and `eHeI_5876` for He I λ 5876 Å and its error
• `HeI_6678` and `eHeI_6678` for He I λ 6678 Å and its error
• `HeII_4686` and `HeII_4686` for He II λ 4686 Å and its error.
• `SII_6716` and `eSII_6716` for [[S II]] λ 6716 Å and its error.
• `SII_6731` and `eSII_6731` for [[[S II]] λ 6731 Å and its error. It is also possible to use the addition of the two [[S II]] lines using `SII_6725` and `SII_6725`.
• `SIII_9069` and `eSIII_9069` for [[S III]] λ 9069 Å and its error.
• `SIII_9532` and `eSIII_9532` for [[S III]] λ 9532 Å and its error. As in the case of [[O III]], if one of the two [[S III]] nebular lines is not introduced, the code assumes its addition taking the theoretical expected ratio.
• `ArIV_4740` and `eArIV_4740` for [ArIV] λ 4740 Å and its error
• `ArIII_7135` and `eArIII_7135` for [ArIII] λ 7135 Å and its error
• `NII_6584` and `eNII_6584` for [NII] λ 6584 Å and its error

The next rows in the input file should correspond to the different values for which the code should derive T_* and log U. The provided emission line fluxes sut be reddening corrected, but it is not necessary that they are relative to Hβ, as only ratios of lines of different ionization are used. Values 0 must be used when no value is found for a line or error.

5.2. *The grids of models*

If the input file is correct, the code will ask for the parameters we want to calculate:

```
(1) Effective temperature and ionization parameter
(2) Photon absorption fraction and ionization parameter

Choose derived parameters:
```

Depending on these parameters, the code will ask for different grid of models and gep,geometry. In the case of T_{eff}:

```
(1) WM-Basic (30-60 kK)
(2) WM-Basic (30-60 kK) and Rauch (80-120 kK) stellar atmospheres
(3) Black body (30-100 kK)

Choose models:
```

The option 1 calculates T_* and log U using WM-Basicsingle star atmospheres from Pauldrach et al., 2001 from 30 to 60 kK, while option 2 extends the range up to 120 kK

using post-AGB stellar atmospheres from Rauch. Option (3) uses black-body spectral energy distributions in the range 30-100 kK.

It is known that the geometry assumed in the models can affect its position on the softness diagrams, so the code will also ask for this in this case:

```
(1) Plane-parallel geometry
(2) Spherical geometry

Choose geometry of the models:
```

On the other hand if we want to calculate the photon absorption factor, f_{abs}, defined as the ratio of ionizing hydrogen photons that do not escape from the nebula, along with $\log U$. This is calculated using cluster model atmospheres from BPASS v.2.1 from Eldridge et al., 2017 assuming an instantaneous burst at 4 Myr, with binaries and an IMF with slope X = -1.35 and an upper mass limit of 300 $M_\odot$. In this case we can choose among models with stellar metallicity identical to that of the gas in each model, or either assume nearly-free metal stars (Z = 10^{-5}. The code also supplies with models with black-body with $T_* = 10^5$ K, equivalent for metal-free stars.

```
(1) BPASS cluster atmospheres, age = 4 Myr, Mup = 300, x = 1.35, w/binaries,
Z* = Zg
(2) BPASS cluster atmospheres, age = 4 Myr, Mup = 300, x = 1.35, w/binaries,
Z* = 1e-5
(3) Black-body, T* = 1e5 K
```

In this case only spherical geometry models are considered.

In Figure 2 it can be seen two examples of the behaviour of the grids of models in one of the softness diagrams as shown in Pérez-Montero et al., 2020 and how the grids of models cover the space of the emission line ratios.

And finally it will ask for the possibility of using interpolation in the final chosen grid:

```
Choose models [0] No interpolated [1] Interpolated:
```

5.3. *Output file and analysis of the results*

As in other versions of HCM, the code will show on the screen the results for each row of the input file along with the ratio of completeness of the task. At the end it will create an output file called as the name of the input file adding `_hcm-teff--ouput.dat`. This file contains all the information for the assumed grids and the results for each one of the rows.

The first column corresponds to the identification of each row while the next ones correspond to the input fluxes of the emission lines. The next two correspond to the assumed O/H, with 0 if no value was introduced. Finally, the four last columns correspond to the calculated T_* (or f_{abs} if the grid 3 was chosen(and $\log U$ with their errors. A value 0 is assigned if no solution is found or the introduced lines are not enough.

In Table 4 I list the mean offsets and the standard deviation of the residuals for the obtained final results as a function of the input emission lines when compared with the input information from the code.

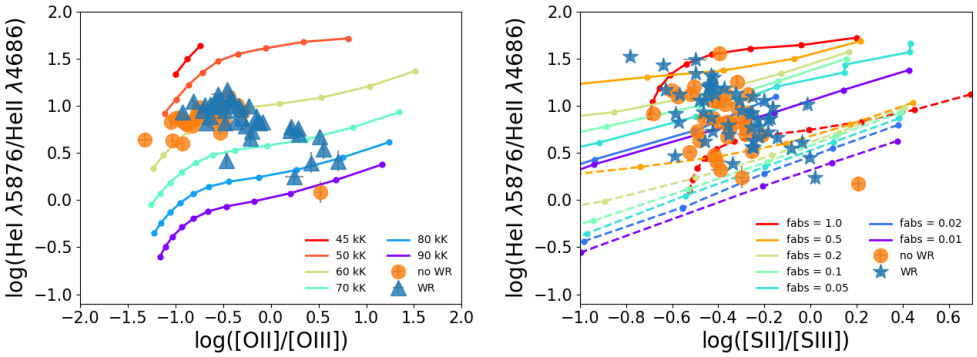

FIGURE 2. Examples of two softness diagrams for the sample of He II-emitters presented in Pérez-Montero et al., 2020. At left [[O II]]/[[O III]] vs He I/He II/ with models at 12+log(O/H) and different values for T_* and, at right, using [[S II]]/[[S III]] and BPASS models assuming different values for f_{abs}. In all sequences values in the lower left part correspond to higher values for log U.

TABLE 4. Mean offsets and standard deviation of the residuals of the resulting properties derived by HCM-TEFF when the model emission lines are used as input as a function of the used emission-line ratios.

Used line ratios	T_* (kK)†		log U^a		log f_{abs}‡		log U^b	
	Mean Δ	σ res.	Mean Δ	σ res.	Mean Δ	σ res.	Mean Δ	σ res.
[[O II]]/[[O III]], [[S II]]/[[S III]], He I/He II	+0.8	2.1	+0.01	0.10	+0.11	0.34	-0.10	0.53
[[O II]]/[[O III]], He I/He II	+0.8	2.1	+0.01	0.10	-0.12	0.27	+0.14	0.47
[[S II]]/[[S III]], He I/He II	+0.6	2.0	+0.03	0.13	-0.19	0.35	+0.10	0.62
[[S II]]/[[O III]], He I/He II	-0.3	5.1	+0.04	0.51	-0.07	0.27	-0.10	0.59
He I/He II	-0.5	6.8	+0.04	0.72	-0.12	0.35	-0.25.	0.63

5.4. *The* HCM-TEFF-IR *version*

A similar code is available for its use based only on mid-IR emission-lines. Although it is planned to merge it with the optical version, so far it workds in an independent module. It is described and applied in Pérez-Montero et al (2023c).

It basically works in the same way as it is described above, using exactly the same grids of models with the exception of those calculated for the derivation of the photon absorption factor. The instructions and steps are the same, but using in the input the following emission-lines, in addition to the `ID` and metallicity entries:

- `ArIII_7m` and `ArIII_7m` for [[Ar III]] at λ 6.98 μm and its error.
- `ArIV_8m` and `ArIV_8m` for [[Ar IV]] at λ 7.90 μm and its error.
- `ArIII_9m` and `ArIII_9m` for [[Ar III]] at λ 8.99 μm and its error.
- `SIV_10m` and `eSIV_10m` for [[S IV]] at λ at 10.5 μm and its error
- `NeII_12m` and `eNeII_12m` for [[Ne II]] λ 12.8 μm and its error.
- `ArV_13m` and `ArV_13m` for [[Ar V]] at λ 13.1 μm and its error.
- `NeV_14m` and `eNeV_14m` for [NeV] λ 14.9 μm and its error.
- `NeIII_15m` and `eNeIII_15m` for [[Ne III]] λ 15.5 μm and its error.
- `SIII_18m` and `eSIII_18m` for [[S III]] λ 18.8 μm and its error.
- `NeV_24m` and `eNeV_24m` for [NeV] λ 24.3 μm and its error.

- `OIV_25m` and `eOIV_25m` for [OIV] λ 25.9 μm and its error.
- `SIII_33m` and `eSIII_33m` for [[S III]] λ at 33.7 μm and its error.
- `OIII_52m` and `OIII_52m` for [[O III]] λ 52 μm and its error.
- `NII_57m` and `eNII_57m` for [[N II]] λ 57 μm and its error
- `OIII_88m` and `eOIII_88m` for [[O III]] λ 88 μm and its error
- `NII_122m` and `eNII_122m` for [[N II]] λ 122 μm and its error.
- `NII_205m` and `eNII_205m` for [[N II]] λ 205 μm and its error.

In this case, the code uses as observables ratios of low- to -high-ionization IR emission-lines that can also be used to construct the corresponding IR softness diagram (e.g. [[Ne II]]/[[Ne III]], [[S III]]/[[S IV]]). This version is less affected by DIG contribution as it involves higher excitation emission-lines. The very high excitation lines, such as [[O IV]] or [[Ne V]] can also be used to discriminate very hard incident SEDs.

Acknowledgements

All versions and articles of the project HII-CHI-MISTRY have been possible to the financial support from the different editions of the Coordinated project of the Spanish Plan Nacional de astronomía y Astrofísica Estallidos de Formación Estelar en Galaxias: Estallidos 5, Estallidos 6, and Estallidos 7.

REFERENCES

Chabrier, G. (2003). The Galactic Disk Mass Function: Reconciliation of the Hubble Space Telescope and Nearby Determinations. *ApJl*, 586:L133–L136.

Eldridge, J. J., Stanway, E. R., Xiao, L., McClelland , L. A. S., Taylor, G., Ng, M., Greis, S. M. L., and Bray, J. C. (2017). Binary Population and Spectral Synthesis Version 2.1: Construction, Observational Verification, and New Results. *PASA*, 34:e058.

Ferland, G. J., Chatzikos, M., Guzmán, F., Lykins, M. L., van Hoof, P. A. M., Williams, R. J. R., Abel, N. P., Badnell, N. R., Keenan, F. P., Porter, R. L., and Stancil, P. C. (2017). The 2017 Release Cloudy. *RMxAA*, 53:385–438.

Fernández-Ontiveros, J. A., Pérez-Montero, E., Vílchez, J. M., Amorín, R., and Spinoglio, L. (2021). Measuring chemical abundances with infrared nebular lines: HII-CHI-MISTRY-IR. *A&A*, 652:A23.

Mollá, M., García-Vargas, M. L., and Bressan, A. (2009). PopStar I: evolutionary synthesis model description. *MNRAS*, 398:451–470.

Pauldrach, A. W. A., Hoffmann, T. L., and Lennon, M. (2001). Radiation-driven winds of hot luminous stars. XIII. A description of NLTE line blocking and blanketing towards realistic models for expanding atmospheres. *A&A*, 375:161–195.

Pérez-Díaz, B., Pérez-Montero, E., Fernández-Ontiveros, J. A., and Vílchez, J. M. (2022). Measuring chemical abundances in AGN from infrared nebular lines: HII-CHI-MISTRY-IR for AGN. *A&A*, 666:A115.

Pérez-Montero, E. (2014). Deriving model-based T_e-consistent chemical abundances in ionized gaseous nebulae. *MNRAS*, 441(3):2663–2675.

Pérez-Montero, E. (2017). Ionized Gaseous Nebulae Abundance Determination from the Direct Method. *PASP*, 129(974):043001.

Pérez-Montero, E. and Amorín, R. (2017). Using photo-ionisation models to derive carbon and oxygen gas-phase abundances in the rest UV. *MNRAS*, 467(2):1287–1293.

Pérez-Montero, E., Amorín, R., Pérez-Díaz, B., Vílchez, J. M., and García-Benito, R. (2023a). Assessing model-based carbon and oxygen abundance derivation from ultraviolet emission lines in AGNs. *MNRAS*, 521(1):1556–1569.

Pérez-Montero, E., Amorín, R., Sánchez Almeida, J., Vílchez, J. M., García-Benito, R., and Kehrig, C. (2021). Extreme emission-line galaxies in SDSS - I. Empirical and model-based calibrations of chemical abundances. *MNRAS*, 504(1):1237–1252.

Pérez-Montero, E. and Contini, T. (2009). The impact of the nitrogen-to-oxygen ratio on ionized nebula diagnostics based on [NII] emission lines. *MNRAS*, 398:949–960.

Pérez-Montero, E. and Díaz, A. I. (2005). A comparative analysis of empirical calibrators for nebular metallicity. *MNRAS*, 361(3):1063–1076.

Pérez-Montero, E., Dors, O. L., Vílchez, J. M., García-Benito, R., Cardaci, M. V., and Hägele, G. F. (2019a). A bayesian-like approach to derive chemical abundances in type-2 active galactic nuclei based on photoionization models. *MNRAS*, 489(2):2652–2668.

Pérez-Montero, E., García-Benito, R., and Vílchez, J. M. (2019b). Revisiting the hardening of the stellar ionizing radiation in galaxy discs. *MNRAS*, 483:3322–3335.

Pérez-Montero, E., Kehrig, C., Vílchez, J. M., García-Benito, R., Duarte Puertas, S., and Iglesias-Páramo, J. (2020). Photon leaking or very hard ionizing radiation? Unveiling the nature of He II-emitters using the softness diagram. *A&A*, 643:A80.

Pérez-Montero, E. and Vílchez, J. M. (2009). On the hardening of the ionizing radiation in HII regions across galactic discs through softness parameters. *MNRAS*, 400:1721–1725.

Pérez-Montero, E., Zinchenko, I. A., Vílchez, J. M., Zurita, A., Florido, E., and Pérez-Díaz, B. (2023b). The softness diagram for MaNGA star-forming regions: diffuse ionized gas contamination or local HOLMES predominance? *A&A*, 669:A88.

Vilchez, J. M. and Pagel, B. E. J. (1988). On the determination of temperatures of ionizing stars in H II regions. *MNRAS*, 231:257–267.