Understanding and reducing statistical uncertainties in nebular abundance determinations

R. Wesson[1,2], D.J. Stock[1,3] & P. Scicluna[1,4]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[2]*European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Santiago, Chile*
[3]*Department of Physics and Astronomy, University of Western Ontario, London, Ontario, Canada, N6K 3K7*
[4]*European Southern Observatory, Karl-Schwarzschildstr. 2, 85748 Garching, Germany*

ABSTRACT

Whenever observations are compared to theories, an estimate of the uncertainties associated with the observations is vital if the comparison is to be meaningful. However, many or even most determinations of temperatures, densities and abundances in photoionised nebulae do not quote the associated uncertainty. Those that do typically propagate the uncertainties using analytical techniques which rely on assumptions that generally do not hold.

Motivated by this issue, we have developed NEAT (Nebular Empirical Analysis Tool), a new code for calculating chemical abundances in photoionised nebulae. The code carries out a standard analysis of lists of emission lines using long-established techniques to estimate the amount of interstellar extinction, calculate representative temperatures and densities, compute ionic abundances from both collisionally excited lines and recombination lines, and finally to estimate total elemental abundances using an ionisation correction scheme. NEAT uses a Monte Carlo technique to robustly propagate uncertainties from line flux measurements through to the derived abundances.

We show that for typical observational data, this approach is superior to analytic estimates of uncertainties. NEAT also accounts for the effect of upward biasing on measurements of lines with low signal to noise, allowing us to accurately determine the effect of this bias on abundance determinations. We find that the effect can result in significant over-estimates of heavy element abundances derived from weak lines, and that taking it into account improves the precision of these abundance determinations. Finally, we investigate the effect of possible uncertainties in R, the ratio of selective to total extinction, on abundance determinations. We find that the uncertainty due to this parameter is negligible compared to the statistical uncertainties due to typical line flux measurement uncertainties.

Key words: ISM: abundances – atomic processes – methods: statistical

1   INTRODUCTION

Abundance determinations from photoionised nebulae play crucial roles in a variety of astrophysical contexts. Nebulae around evolved stars, e.g. Planetary Nebulae (PNe) or Wolf-Rayet (WR) ejecta nebulae provide constraints on theories of stellar nucleosynthesis and evolution (e.g. Karakas et al. 2009; Magrini et al. 2011; Maeder 1992; Stock et al. 2011). Such data is invaluable as constraints on stellar yields; the *inputs* of galactic chemical evolution models. While Galactic and extragalactic H II regions provide insights into the current composition of the ISM and therefore are vital constraints for the *output* of galactic chemical evolution models (e.g. Pagel 1997; Matteucci 2003).

Such studies have a very long history (Huggins & Miller 1864) **(is this gratuitously ancient enough?)**, but some major questions remain unanswered. One long standing issue is the sometimes sizable discrepancy between abundances derived from collisionally excited lines (CELs) and those derived from recombination lines (RLs) (Wesson et al. 2005; Liu et al. (2006); García-Rojas & Esteban 2007; Tsamis et al. 2008).

A meaningful comparison of observations with models is only possible if the uncertainties on the observations can be meaningfully estimated. However, such a meaningful estimate is not trivial to determine. The final uncertainty will be due to both statistical uncertainty (ultimately deriving from

the inherent uncertainty on the original line flux measurements) and systematic uncertainty (for example, the choice of reddening law or atomic data set). Sources of systematic uncertainty may be unknown and even unquantifiable.

In this paper we present a new code for calculating chemical abundances in photoionised nebulae, which can also robustly estimate statistical uncertainties using a Monte Carlo approach, so called because of its reliance on probabilities. This method is inherently superior to analytic methods of uncertainty propagation. and can easily account for the effect of non-Gaussian flux uncertainties, such as those found by Rola & Pelat (1994) in the case of flux measurements with low signal-to-noise ratios.

We use our code to reanalyse several published line lists for which uncertainties are available, and we show firstly that analytic methods can significantly underestimate the true uncertainties on derived quantities. Typically, uncertainties on derived abundances are better characterised by log-normal distributions than by normal distributions, with most quantities being potentially more subject to upward bias than to downward bias. In some cases, bimodal probability distributions emerge.

Secondly, as the Monte Carlo approach can trivially handle any quantifiable distribution of line flux uncertainties, we have designed the code to account for the well known upward bias in measurements of weak lines, and associated log-normal distribution of measurement uncertainties (Rola & Pelat (1994)). The bias is well known but its actual effect on abundance determinations has not previously been quantified. We show, as expected, that abundances determined from weak lines are systematically overestimated. Moreover, we show that accounting for the effect leads to a significant improvement in the precision of abundance determinations from weak lines.

Finally, we investigate the effect of an assumed statistical uncertainty in R, the ratio of selective to total extinction. We show that the extra uncertainty introduced into abundance determinations by taking this parameter to be 3.1 $\pm$ 0.15 instead of a fixed value of 3.1 is small compared to the uncertainties arising from line flux measurements, even when the extinction is quite large and the line fluxes are well measured.

The words "uncertainty" and "error" are often used synonymously. In this article we maintain a distinction in meaning between the terms: "uncertainty" refers to the limiting accuracy of the knowledge of a quantity, while "error" refers to an actual mistake.

2 STATISTICAL UNCERTAINTIES

Uncertainties in observed quantities can be propagated into the uncertainty on derived parameters in a number of different ways, the two most common of which are the traditional analytical technique based on systems of partial derivatives and simplifying assumptions that allow one to apply Taylor expansions, and the 'Monte Carlo' method which is a brute-force iterative method that exploits the wealth of computational power now readily available by building on knowledge of the uncertainty in the original observations.

The analytical approach is as follows: if one has measured a quantity x with some uncertainty σx, and wishes to calculate the uncertainty in a quantity F given that $F = f(x)$, then the uncertainty on F can be computed via the relation

$$\frac{\sigma F}{\sigma x} \simeq \frac{\partial f}{\partial x} \qquad (1)$$

However, in general one may not be able to compute this partial differential exactly, and in these cases, provided that

$$\frac{\partial f}{\partial x}|_{x=x_1} \ll f(x_1) \qquad (2)$$

it is possible to use a first order Taylor expansion to approximate this derivative. If one had a third value, $H = h(f, g)$ where f and g are both functions of other variables and are statistically independent of one another, it would be necessary to find the total derivative dH such that

$$dh^2 = \left(\frac{\partial h}{\partial f}\right)^2 df^2 + \left(\frac{\partial h}{\partial g}\right)^2 dg^2 \qquad (3)$$

and it thus follows that

$$\sigma_H^2 = \left(\frac{\partial h}{\partial f}\sigma f\right)^2 + \left(\frac{\partial h}{\partial g}\sigma g\right)^2 \qquad (4)$$

This expression can be generalised to any number of variables, and gives rise to the usual quadrature formulae for many simple functions of x. We highlight three key aspects of this approach:

(i) Given the number of formulae through which the original line flux data must be put before an abundance can be determined, and the wide variety of functions applied, the equations necessary to propagate uncertainties in this way can become extremely complex;

(ii) the approach implicitly assumes that all the input and output uncertainties at each step are normally distributed;

(iii) the Taylor expansion requires that the uncertainties be small relative to the quantities.

The first point is a matter of convenience but none the less one which discourages many authors from even attempting to propagate uncertainties all the way into the final quantities. The second and third points are clearly violated in many or most real astrophysical observations, by virtue of which the uncertainties estimated using analytical techniques are liable not to reflect the true uncertainties.

The Monte Carlo method, on the other hand, exploits the fact that if an observation of a quantity is drawn from a distribution X, with mean x and variance σ_x, and if one knows (or can make sensible assumptions about) the shape of this distribution, then a random-number generator can be used to repeatedly draw values from X, creating a random sample from it. Using the above example of $F = f(x)$, if one wanted to examine the uncertainty in F, the operation f(x) could be performed upon every value in the aforementioned sample to produce a sample of the distribution from which F is drawn, which can then be parameterised to estimate the type, mean and standard deviation of this distribution. This process can be repeated ad infinitum for any number or combination of functions of F, x, or any other variable derived in the same way to propagate the uncertainties on the quantities, irrespective of the size of the uncertainties and any statistical interdependence of the variables. This

approach thus has the following advantages over the analytical approach:

(i) It is inherently very simple;

(ii) it requires no assumption about the particular distribution of uncertainties at any stage;

(iii) it does not require relative uncertainties to be small

The Monte Carlo approach is thus inherently robust when applied to real astrophysical data, in a way that the analytical approach is not. The only limitation is then the time taken to run the calculation enough times to sample well the statistics of the output distributions.

3 NEAT: NEBULAR EMPIRICAL ABUNDANCE TOOL

We have developed a new code, NEAT (Nebular Empirical Analysis Tool), to quickly carry out a thorough analysis of emission line spectra, and to propagate statistical uncertainties using the Monte Carlo technique described above. In this section we describe the code and how it works.

3.1 Input

NEAT incorporates elements of several previous codes, most importantly the EQUIB code, also developed at UCL, for solving the equations of statistical equilibrium in multi-level atoms. All the source code, documentation, atomic data and example line lists are freely available at `https://github.com/rwesson/NEAT`. The code is designed to be as simple as possible from the user's perspective, and our aim is that the user should be able to simply pass the code a line list and get out abundances determined robustly. The code requires no external libraries and should compile without problems on any Unix system. We have also verified that it compiles and runs on Windows systems, should the user be restricted to such an OS.

Hydrogen recombination data from Storey & Hummer (1995) is used throughout. All atomic data for heavier elements is stored externally in plain text files, so that the user can easily change the atomic data being used without needing to recompile the code. We provide three sets of atomic data for collisionally excited lines with the code - a compilation from a variety of sources compiled on an ad hoc basis, CHIANTI 5.2 (Landi et al. 2006), and CHIANTI 6.0 (Dere et al. 2009; **?**). In all the analyses presented in this paper, we used atomic data from CHIANTI 5.2, with the exception of data for O^{2+} for which a documented error exists in the CHIANTI 5.2 data (Kisielius et al. 2009), and S^{3+}, which we believe is affected by a similar error. Data for these two ions was instead taken from (cite) and (cite) respectively. For recombination lines, we use data from the sources given in Table 3.1.

NEAT requires as input a plain text list of rest wavelengths, line fluxes, and uncertainties. The user can then select the number of iterations of the code to run. If the number of iterations is one, the code performs a standard empirical analysis on the line list, as described below, and does not calculate any uncertainties. If the number of iterations is more than one, then the code first randomises the

Ion	Recomination data source
C^{2+}	
C^{3+}	
N^{2+}	
N^{3+}	
O^{2+}	
Ne^{2+}	

Table 1. Atomic data used in NEAT for recombination lines

line list, as described below. The standard analysis is then carried out on the randomised line list.

By randomising and analysing the line list many times, it is possible to build up an accurate picture of the true distribution of statistical uncertainties associated with the chemical abundances and empirical diagnostics resulting from the line flux uncertainties. NEAT collates all of the results from each iteration, and generates histograms showing visually the uncertainty distributions on the output parameters.

3.1.1 Line flux randomisation

The code randomises the line fluxes by assuming that they come from one of three distributions, depending on the measured signal to noise. The three cases are:

SNR>6.0: the line flux is assumed to be drawn from a Gaussian distribution. In this case, if the given line flux measurement is F and the given measurement uncertainty is σ, then in each iteration of the code the line flux is calculated using

$$F_i = F + R \times \sigma \qquad (5)$$

where R is a random number drawn from a Gaussian distribution with a mean of zero and a standard deviation of unity.

1.0<SNR<6.0: Rola & Pelat (1994, hereafter RP94) observed that measurements of weak lines (SNR<6) are strongly biased upwards, and the uncertainty on such lines cannot be well represented by a normal distribution. We refer to this effect henceforth as the RP effect. Taking the RP94 effect into account is straightforward with a Monte Carlo approach, requiring only that a log-normal distribution with appropriate mean and standard deviation is used for weak lines, instead of the normal distribution used for strong lines.

For lines with $F/\sigma < 6$, we randomise the line fluxes using the following procedure: we first determine the log-normal distribution appropriate to the measured signal-to-noise ratio using the following equations, which are derived from fits to the numbers in Table 6 of RP94:
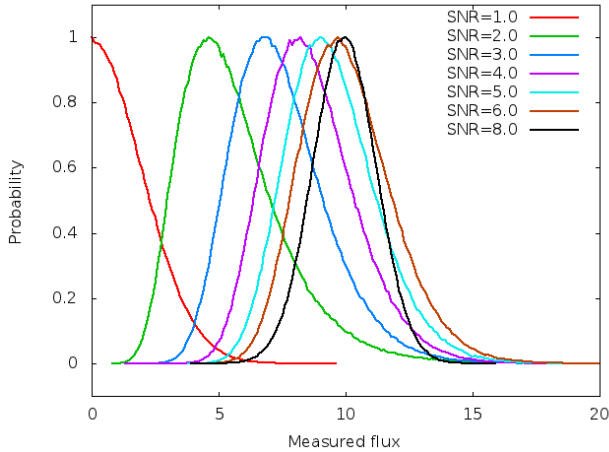
$$\mu = \frac{0.0765957}{snr^2} + \frac{1.86037}{snr} - 0.309695 \qquad (6)$$

$$\sigma = \frac{-1.11329}{snr^3} + \frac{1.8542}{snr^2} - \frac{0.288222}{snr} + 0.18018 \qquad (7)$$

Then, the line flux is found using the following equation:

$$F_i = \frac{F}{e^{(R \times \sigma + \mu)}} \qquad (8)$$

where R is a random number from a gaussian distribution as before. The result of this procedure is that weaker

Figure 1. The behaviour of NEAT's line flux randomization as a function of SNR, for a fixed measured flux of 10.0. At high SNR (i.e. >6), normal distributions apply. At lower SNR, the effect described by Rola & Pelat (1994) results in log-normal distributions which are skewed towards lower values than the measured intensity. If SNR⩽1 then the code assumes that the quoted flux is a 5σ upper limit

lines are drawn from log-normal distributions that peak below the observed value, with the effect increasing as SNR decreases. The statistical behaviour is as described by RP94.

SNR<1.0: If the quoted uncertainty is larger than the acual line flux, the code assumes that the quote flux represents a 5σ upper limit, and thus draws the randomized line flux from a folded gaussian distribution, with $\mu=0$ and $\sigma=0.2\times$F. The line flux is given by

$$F_i = abs(R) \times 0.2F \qquad (9)$$

Figure 1 shows the various possibilities that arise depending on the measured SNR. The distributions plotted are from 10^6 runs of NEAT in which 7 artificial line fluxes were randomized. Each of the lines had a measured flux of 10.0, and the quoted uncertainties represented SNRs of 1.0, 2.0, 3.0, 4.0, 5.0, 6.0 and 8.0. The figures shows how the normal distribution appropriate at high SNR is replaced by a log-normal distribution increasingly skewed towards lower values as SNR decreases.

3.1.2   Random number algorithm

This process relies on the FORTRAN pseudo-random number generator, seeded using the system clock and the time of the code's exection. The uniformly distributed numbers thus generated are then converted into a gaussian distribution using an algorithm based on the ratio of uniforms method of Kinderman & Monahan (1977), available from `netlib.org`. We tested the performance of this method by running the code 1 000 000 times, and plotting the distribution of fluxes obtained for an arbitrarily selected line. We then fitted a gaussian function to this distribution. We found that the recovered mean was within 0.008% of the specified value, while the recovered standard deviation was within 0.13% of the specified value. Figure 2 shows the histogram of generated values with the required gaussian distribution overplotted.
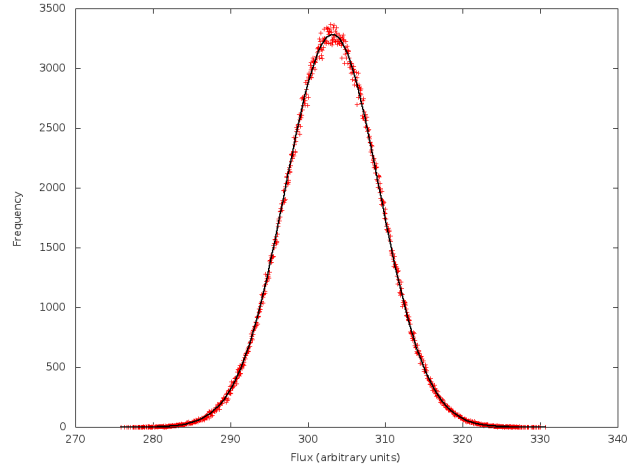
Figure 2. A distribution of values produced by 1 000 000 runs of the random number generator, with the target gaussian distribution overplotted.

We thus consider that the random number generator in the code provides a reliably random Gaussian distribution.

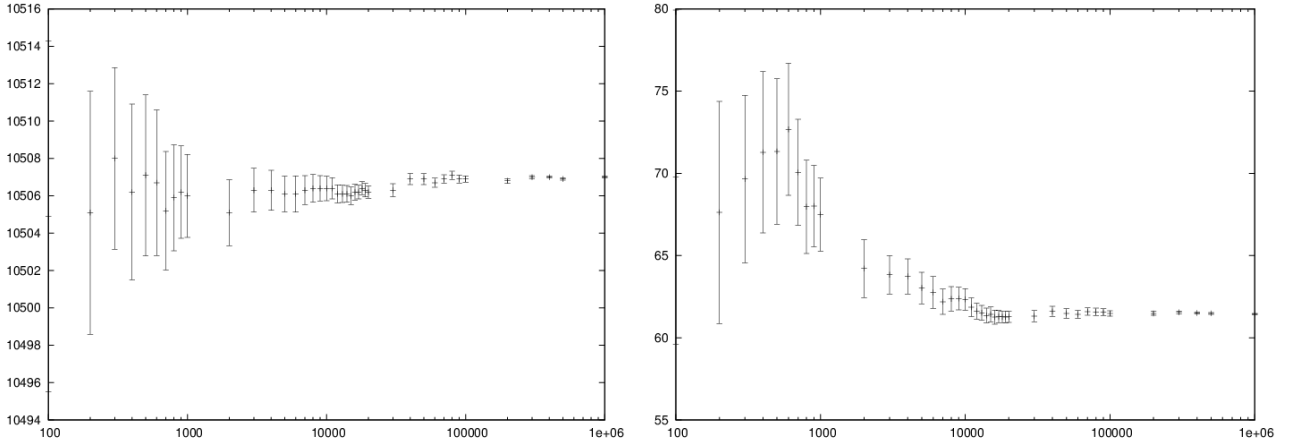### 3.1.3   Sampling uncertainty

The Monte Carlo approach relies on carrying out the analysis enough times to adequately sample the probability distributions of the derived quantities. To determine what number of iterations suffices for this purpose, we ran the code 1,000,000 times, using emission line fluxes measured for IC1747 by Wesson et al. (2005). We then considered subsets of the output from this run.

To quantify the sampling uncertainty, we fitted a gaussian to the observed probability distribution of the measured [O III] temperatures, for each subset of iterations. We chose this quantity as its actual uncertainty distribution should closely approximate a gaussian. The uncertainty on the gaussian fit is thus a measure of how well the output distribution was sampled, for a given number of iterations.

In Figure 3 we show how the uncertainty of the fitted μ and σ vary with the total number of iterations. We find that the precision of the fit improves indefinitely with increasing number of iterations up to the limit of our investigation. However, we find that approximately 10,000 iterations provides a sensible trade-off - more iterations than this gives little absolute reduction in the sampling uncertainty. We therefore carry out 10,000 iterations on each line list used in our analyses here.

3.2   Interstellar extinction

The first step of any abundance analysis is a correction for interstellar extinction. The amount of extinction is determined by NEAT from the ratios of hydrogen Balmer lines, in an iterative procedure: the extinction is first calculated assuming intrinsic Hα, Hβ and Hγ line ratios for a temperature of $10\,000$ K and a density of 1000 cm^{-3}. The line list is dereddened using the value of c(Hβ) thus obtained, and temperatures and densities calcaulted as described below. The intrinsic Balmer line ratios are then recalculated

Figure 3. The dependence of the sampling uncertainty on the number of iterations. Sampling uncertainty is quantified as described in the text. 10,000 iterations provides a sensible balance of acceptable processing time and sampling uncertainty.

at the appropriate temperature and density, and the line list dereddened using this revised value.

The user can select the particular extinction law to be used. Five extinction laws are currently available: the Galactic extinction curves of Howarth (1983), Fitzpatrick & Massa (1990) and Cardelli et al. (1989), the Large Magellanic Cloud law of Howarth (1983), and the Small Magellanic Cloud law of Prevot et al. (1984). Adding further extinction laws would be straightforward, should the user wish to do so.

In section 7 we investigate the effect on derived quantities of the likely uncertainty in R, the ratio of selective to total extinction.

3.3 Temperatures and densities

Temperatures and densities are calculated using traditional collisionally excited line diagnostics. For the purposes of subsequent abundance calculations, the nebula is divided into three "zones", of low, medium and high excitation. In each zone, temperatures and densities are calculated iteratively and weighted according to the reliability of each diagnostic. Table 3.3 shows the diagnostics used and the weighting given in each zone.

The scheme to calculate the densities is iterative, and proceeds as follows:

(i) A temperature of 10,000K is initially assumed, and the density is then calculated from the line ratios relevant to the zone.

(ii) The temperature is then calculated from the temperature diagnostic line ratios, using the derived density.

(iii) The density is recalculated using the appropriately weighted average of the temperature diagnostics

(iv) The temperature is recalculated using this density.

This iterative procedure is carried out successively for low, medium and high ionisation zones, and in each case if no diagnostics are available, the temperature and/or density will be taken to be those derived for the previous zone.

Low ionisation zone		
Diagnostic	Lines	Weight
[O II] density	$\lambda3727/\lambda3729$	1
[S II] density	$\lambda6717/\lambda6731$	1
[N II] temperature	$\frac{\lambda6548+\lambda6584}{\lambda5754}$	5
[S II] temperature	$\frac{\lambda6717+\lambda6731}{\lambda4068+\lambda4076}$	1
[O II] temperature	$\frac{\lambda7319,20+\lambda7330,31}{\lambda3726+\lambda3729}$	1
[O I] temperature	$\frac{\lambda6363+\lambda6300}{\lambda5577}$	1
[C I] temperature	$\frac{\lambda9850+\lambda9824}{\lambda8727}$	1

Medium ionisation zone		
Diagnostic	Lines	Weight
[Cl III] density	$\lambda5517/\lambda5537$	1
[Ar IV] density	$\lambda4711/\lambda4740$	1
[C III] density	$\lambda1907/\lambda1909$	1
[O III] density	$\lambda52\mu m/\lambda88\mu m$	0
[Ar III] density	$\lambda9\mu m/\lambda21.8\mu m$	0
[S III] density	$\lambda18.7\mu m/\lambda33.8\mu m$	0
[Ne III] density	$\lambda15.5\mu m/\lambda36.0\mu m$	0
[O III] temperature	$\frac{\lambda4959+\lambda5007}{\lambda4363}$	4
[Ar III] temperature	$7135 + 7751/5192$	2
[Ne III] temperature	$3868 + 3967/3342$	2
[S III] temperature	$9069 + 9531/6312$	1
[Ne III] temperature	$3868 + 3967/15.5\mu m$	0
[O III] temperature	$4959 + 5007/52$	0

High ionisation zone		
Diagnostic	Lines	Weight
[Ne IV] density	$2423/2425$	1
[Ar V] temperature	$6435 + 7005/4625$	1
[Ne V] temperature	$3426 + 3345/2975$	1

Table 2. Diagnostics used in the calculation of physical conditions.

3.4 Ionic abundances

Ionic abundances are calculated from collisionally excited lines using the temperature and density appropriate to their ionisation potential. Where several lines from a given ion are

present, the ionic abundance adopted is found by averaging the abundances from each ion, weighting according to the observed intensity of the line.

Recombination lines are also used to derive ionic abundances for helium and heavier elements. In deep spectra, many more recombination lines may be available than collisionally excited lines. The code first assigns IDs to each line based on their rest wavelengths, and then calculates the ionic abundance from the line intensity using the atomic data listed in Table 3.1. Then, to determine the ionic abundance to adopt, it first derives an ionic abundance for each individual multiplet from the multiplet's co-added intensity, and then averages the abundances derived for each multiplet to obtain the ionic abundance used in subsequent calculations.

3.5   Total elemental abundances

Generally, not all ionisation stages of an ion that are actually present in a nebula will be detected, due to limited wavelength coverage and sensitivity. Total elemental abundances must be estimated using ionisation correction schemes, which are derived from photoionisation models, or similarities in ionisation potentials, or a combination of the two.

The code currently includes the ICF scheme of Kingsburgh & Barlow (1994). We plan to incorporate further ICFs, and in a forthcoming paper we will compare the magnitude of the systematic uncertainties arising from the choice of ICF with the statistical uncertainties.

4   COMPARING ANALYTIC AND MONTE CARLO UNCERTAINTIES

To demonstrate how traditional methods of uncertainty propagation do not properly reflect the true uncertainties on the derived chemical abundances, we present here a reanalysis using NEAT of observations of the nebula surrounding the Large Magellanic Cloud Wolf-Rayet star WR18. These observations were presented by Stock et al. (2011) and analysed using an earlier version of NEAT. In this section we calculate the abundances and uncertainties using both the current version of NEAT, which accounts for non-Gaussian line flux uncertainties, and using analytical techniques.

We chose this object as it is distant and faint, and the line measurements are thus necessarily measured with relatively low signal to noise ratios, and the effects of the assumptions in the analytical method are most apparent. The results of this analysis are presented in Table **??**.

5   THE NATURE OF ACTUAL UNCERTAINTIES

General comments on what typically emerges - approximately gaussian for quantities derived from simple functions of well-measured inputs, mostly lognormal for most things, sometimes crazy.

6   THE RP EFFECT

As discussed earlier, NEAT accounts for the RP effect, in which fluxes measured from lines with SNR<6 are generally overestimated, with the magnitude of the effect increasing as SNR→1. In this section we show the importance of this effect and demonstrate that flawed results will inevitably result if the effect is ignored.

To investigate the magnitude of this effect, we reanalysed two line lists - spectra of NGC 6543, the Cat's Eye Nebula, presented by Wesson & Liu (2004), and spectra of the Orion Nebula presented by Esteban et al. (2004). In both cases we ran two instances of NEAT, one in which the RP effect was ignored, and all line flux uncertainties quoted in the two papers were assumed to represent gaussian probability distributions, and the second in which the RP effect was accounted for as described in Section 3.1.1.

One important and as yet unresolved issue in nebular abundance studies is the so-called abundance discrepancy problem (see for example Liu (2006) for a recent review). The RP effect can cause errors in the assessment of the magnitude of the discrepancy; recombination lines of heavy elements are much weaker than the collisionally excited lines of the same species, and are thus measured with lower signal to noise ratios. In almost any real astronomical spectra, regardless of the number of lines detected, the weakest lines measured will be subject to the RP effect, and for deep spectra of photoionised nebulae, the weakest lines will almost all be recombination lines. Thus, recombination line abundance measurements may be subject to an upward bias that collisinally excited line abundances are largely free from.

Figure 4 shows that in these two nebulae, the RP effect indeed has no effect on abundances of nitrogen, oxygen and neon derived from collisionally excited lines, but in all cases leads to an overestimate of the abundances derived from the recombination lines of these elements. Furthermore, it turns out that properly accounting for the non-Gaussian nature of the uncertainties on weak lines leads to a significant improvement in the precision of abundances determined from them; we fit the resulting uncertainty histograms with gaussian functions, and find that accounting for the RP effect reduces the standard deviation on the measured abundances by 30-40%. Table 3 summarises the mean and standard deviation of the final abundance determinations from recombination lines.

In other shallower spectra, it may often be the case that the auroral lines [N II] $\lambda 5754$ and [O III] $\lambda 4363$ are measured with low enough SNR that they become subject to the RP effect. In this case, the derived temperatures would be overestimated, and collisionally excited line abundances underestimated. To investigate this effect, we analysed the line list of H1013, an extragalactic H II region in the spiral galaxy M101, presented by Esteban et al. (2009). In this nebula, the [O III] line at 4363Å is detected with an SNR of only 2.7.

Figure 5 shows that significant systematic uncertainties are produces when the RP effect on temperature diagnostic lines is ignored. When the effect is neglected, we determine an [O III] temperature of 7480±610 K, in very close agreement with the value of 7370±630 K reported by Esteban et al. (2009). However, when we account for the RP effect, we find a value of 6840±390 K. Similarly for abundances,

Element	$10^4 \times$ X/H (RP ignored)	$10^4 \times$ X/H (RP included)	$\frac{\sigma(RP\,included)}{\sigma(RP\,ignored)}$
		Orion	
N	2.74 ± 0.16	2.37 ± 0.11	0.69
O	7.04 ± 0.63	6.63 ± 0.43	0.69
Ne	1.52 ± 0.26	1.20 ± 0.17	0.66
		NGC 6543	
N	6.93 ± 0.56	5.83 ± 0.37	0.66
O	14.94 ± 1.15	12.48 ± 0.74	0.65
Ne	4.17 ± 0.94	3.23 ± 0.56	0.60

Table 3. Results of heavy element abundance determinations from recombination lines, using the published line lists of Wesson & Liu (2004) and Esteban et al. (2004) for the Orion Nebula.

considering O^{2+}/H^+, we find that by neglecting the RP effect we obtain a value of 7.95 ± 0.15 (on a logarithmic scale where N(H)=12), close to the value of 8.05 ± 0.12 obtained by Esteban et al. (2009). Accounting for the effect yields a value of 8.16 ± 0.12.

We re-emphasise, therefore, that neglecting this effect results in incorrect abundances. Accounting for it improves both the accuracy and precision of the abundances determined.

7 INTERSTELLAR EXTINCTION

NEAT allows a robust propagation of uncertainties from line flux measurements into derived quantities. It is also possible to investigate the effect of uncertainties arising at different stages of the process. In this section, we consider the effect of the uncertainty in R, the ratio of total to selective extinction given by

$$R = \frac{A(V)}{E(B-V)} \tag{10}$$

It is well known that R varies along different sightlines (eg Valencic et al. (2004), Larson & Whittet (2005)), but estimating its value for particular objects is generally impractical and instead, it is commonly assumed to equal 3.1. We investigate the effect of an uncertainty in this value by comparing analyses in which R is fixed to be 3.1, and in which R is drawn from a Gaussian distribution with μ=3.1 and σ=0.15. For this investigation we used the R-dependent extinction law parametrization of Cardelli et al. (1989). We took the uncertainty as the largest value that we found quoted in the literature for the diffuse ISM (Larson & Whittet 2005).

We analysed the emission line measurements of NGC 7026 presented by Wesson et al. (2005), including the line flux uncertainties which were not published in that paper. We chose this object as it is significantly reddened (c(Hβ)=1.0), and has many very well detected lines in its spectrum (142 lines measured, 120 with SNR>3 and 65 with SNR>10). This combination should maximise the effect of an uncertainty on R, relative to the effect of the line flux measurement uncertainties.

We find that including the effect of an uncertainty in R has a noticeable effect on the probability distribution of dereddened line fluxes. However, in the conversion from line fluxes into physical quantities, the statistical uncertainties arising from line flux measurements completely dominate, and the probability distributions are statistically identical whether R is assumed to be fixed or allowed to vary. This

result is shown in Figure 6, where we plot the probability distributions for the [Ne III] 3868Å dereddened line flux, and the abundance derived from it. The effect on the derived abundance of the assumed uncertainty on R is negligible.

8 DISCUSSION

Motivated by the necessity of properly understanding sources of uncertainties in analyses of photoionised nebulae, we have presented a new code for calculating chemical abundances in photoionised nebulae, which also robustly calculates the statistical uncertainties on the abundances determined. The code is freely available and we welcome bug reports and feature requests.

We have demonstrated that analytic methods of uncertainty propagation rely on assumptions that typically do not hold for real astronomical data, and therefore we have adopted a Monte Carlo approach. We have shown that the analytic approach does not give a good estimate of the true statistical uncertainties.

We have provided in the code a means for accounting for the well known upward bias in flux measurements of weak lines, and we show that doing so results in significantly improved precision in abundance determinations. This effect should not be ignored in any analyses of emission line spectra; in almost any data set, the weakest lines will be subject to the effect, and incorrect results will be obtained if the effect is not properly accounted for. If, as is commonly the case, recombination lines make up the majority of the weak lines, then systematic misunderstandings may arise if the effect is neglected. However, in the examples that we have analysed, this effect is certainly too small to account for the magnitude of the discrepancy typically found between RL and CEL abundance determinations.

Finally we have shown that possible uncertainties in the value of R have a negligibly small effect on the uncertainties on the final quantities. This touches upon a further application for the code that we plan to pursue in a forthcoming paper; by varying the *systematic* approach being used – for example, using different atomic data sets, reddening laws or ionisation correction schemes – we can compare the magnitude of the resulting systematic effects with the true statistical uncertainties, and thus determine realistically which systematic effects are the most important.
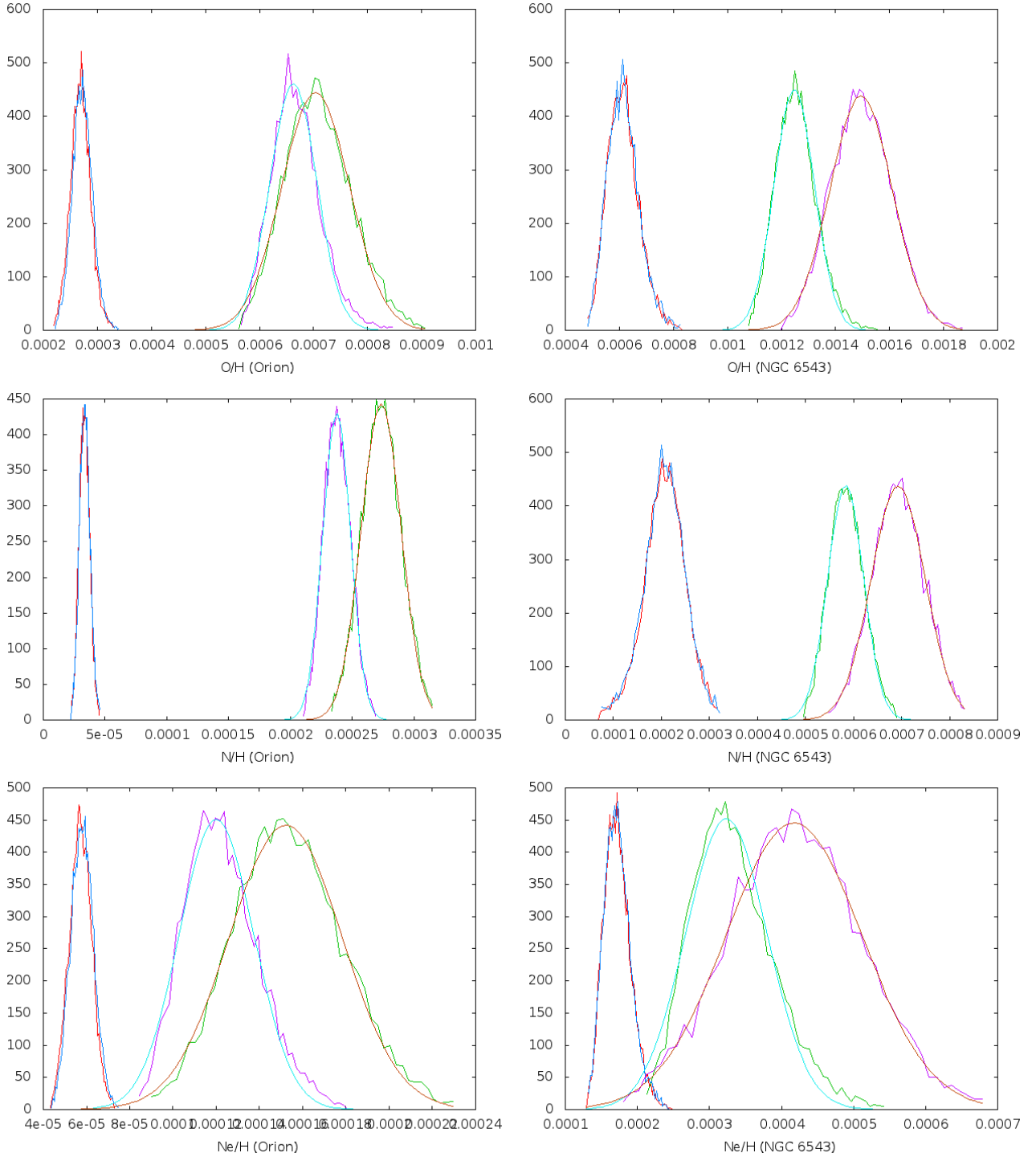
Figure 4. Figures comparing abundance determinations from weak lines carried out by incorrectly assuming normally distributed uncertainties with those in which weak line uncertainties are characterised following RP94

REFERENCES

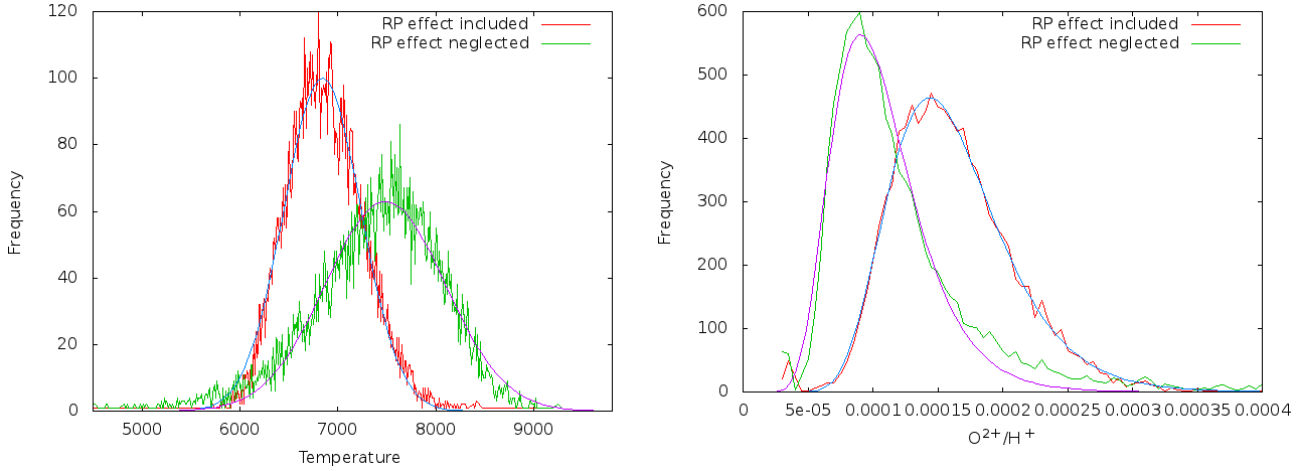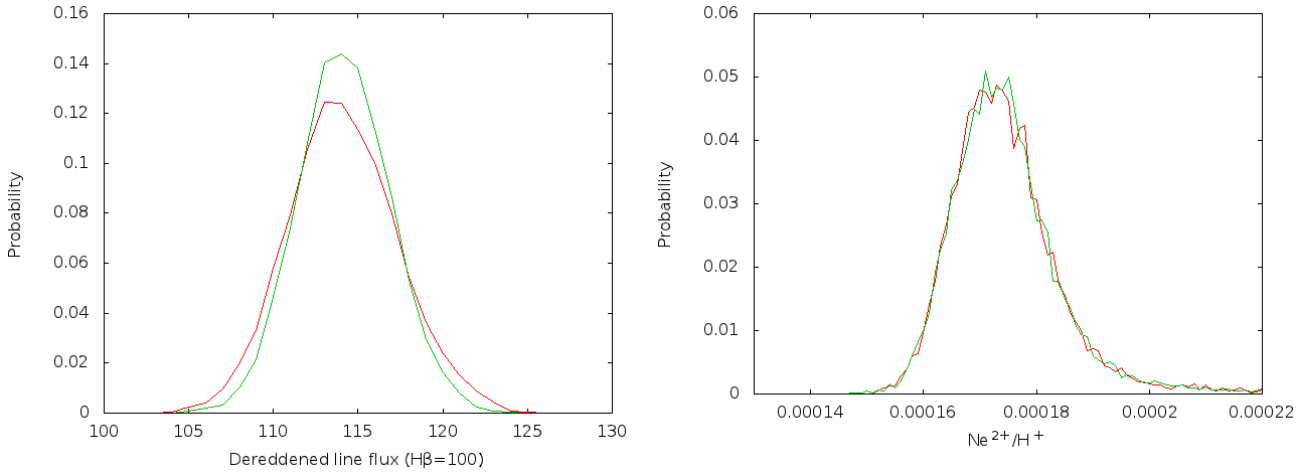Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

Figure 5. The RP effect when CEL temperature diagnostic lines are measured with low SNR. The figure shows a reanalysis of data for the H II region H1013, showing the RP effect on the derived [O III] temperature (l) and O^{2+}/H^+ abundance (r).

Figure 6. The effect of an assumed uncertainty in R, the ratio of selective to total extinction, on the dereddened line flux of the [Ne III] line at 3868Å (l) and derived Ne^{2+}/H^+ abundance (r).

Dere K. P., Landi E., Young P. R., Del Zanna G., Landini M., Mason H. E., 2009, A&A, 498, 915

Esteban C., Bresolin F., Peimbert M., García-Rojas J., Peimbert A., Mesa-Delgado A., 2009, ApJ, 700, 654

Esteban C., Peimbert M., García-Rojas J., Ruiz M. T., Peimbert A., Rodríguez M., 2004, MNRAS, 355, 229

Fitzpatrick E. L., Massa D., 1990, ApJS, 72, 163

García-Rojas J., Esteban C., 2007, ApJ, 670, 457

Howarth I. D., 1983, MNRAS, 203, 301

Huggins W., Miller W. A., 1864, Royal Society of London Philosophical Transactions Series I, 154, 437

Karakas A. I., van Raai M. A., Lugaro M., Sterling N. C., Dinerstein H. L., 2009, ApJ, 690, 1130

Kinderman A. J., Monahan J. F., 1977, ACM Trans. Math. Softw., 3, 257

Kingsburgh R. L., Barlow M. J., 1994, MNRAS, 271, 257

Kisielius R., Storey P. J., Ferland G. J., Keenan F. P., 2009, MNRAS, 397, 903

Landi E., Del Zanna G., Young P. R., Dere K. P., Mason H. E., Landini M., 2006, ApJS, 162, 261

Larson K. A., Whittet D. C. B., 2005, ApJ, 623, 897

Liu X.-W., 2006, in M. J. Barlow & R. H. Méndez ed., Planetary Nebulae in our Galaxy and Beyond Vol. 234 of IAU Symposium, Optical recombination lines as probes of conditions in planetary nebulae. pp 219–226

Liu X.-W., Barlow M. J., Zhang Y., Bastin R. J., Storey P. J., 2006, MNRAS, 368, 1959

Maeder A., 1992, A&A, 264, 105

Magrini L., Stanghellini L., Goncalves D. R., 2011, ArXiv e-prints

Matteucci F., 2003, The Chemical Evolution of the Galaxy

Pagel B. E. J., 1997, Nucleosynthesis and Chemical Evolution of Galaxies

Prevot M. L., Lequeux J., Prevot L., Maurice E., Rocca-Volmerange B., 1984, A&A, 132, 389

Rola C., Pelat D., 1994, A&A, 287, 676

Stock D. J., Barlow M. J., Wesson R., 2011, MNRAS, p. 1754

Storey P. J., Hummer D. G., 1995, MNRAS, 272, 41

Tsamis Y. G., Walsh J. R., Péquignot D., Barlow M. J.,

Danziger I. J., Liu X.-W., 2008, MNRAS, 386, 22

Valencic L. A., Clayton G. C., Gordon K. D., 2004, ApJ, 616, 912

Wesson R., Liu X.-W., 2004, MNRAS, 351, 1026

Wesson R., Liu X.-W., Barlow M. J., 2005, MNRAS, 362, 424