International Workshop on Stellar Spectral Libraries
ASI Conference Series, 2014, Vol. 11, pp 67 – 75
Editors: H. P. Singh, P. Prugniel and I. Vauglin

ALiCCE: Atomic Lines Calibration using the Cross-Entropy algorithm

L. P. Martins[1]*, R. Vitoriano[1], P. Coelho[1,2] and A. Caproni[1]
[1]*NAT - Universidade Cruzeiro do Sul, Rua Galvão Bueno, 868, São Paulo, SP, Brazil*
[2]*IAG - Universidade de São Paulo, R. do Matão, 1226, 05508-090, São Paulo, SP, Brazil*

Abstract. Atomic line opacities play a crucial role in stellar astrophysics. However, very few parameters of atomic transitions were actually measured in the laboratory and most of the lines were calculated with low precision atomic parameters. Only a small part of the lines were calibrated empirically. These calibrations are generally done manually for a small subset of lines. For the purpose of computing a complete stellar grid with a good range of atmospheric parameters, this situation is rather limiting. We have implemented an innovative method to perform this calibration in a more robust and efficient way called ALiCCE: Atomic Lines Calibration using the Cross-Entropy algorithm. Here we present the adaptation of the method to the calibration of atomic lines used by spectral synthesis codes. This is done using synthetic models of high quality. We show here that the method is extremely efficient for calibration of line lists and can be used to do that using high quality spectra of real stars.

Keywords : atomic data – methods: numerical – stars: general

1. Introduction

Atomic line opacities play a crucial role in stellar astrophysics. However, we still lack a complete, accurate and precise list of atomic transitions. In general, one half of the discernible lines in observed stellar spectra are missing from the lists of lines with good wavelengths (Kurucz 2011). Except for the simplest species, it is impossible to generate accurate energy levels or wavelengths theoretically. In principle, they must be measured in the laboratory.

*email: lucimara.martins@cruzeirodosul.edu.br

As very few parameters were measured in the laboratory, to compute a theoretical stellar spectra with good spectrophotometry it is necessary to include the so-called "predicted lines": lines where either one or both energy levels of the transition were predicted from quantum mechanics calculations (Kurucz 1992). The predicted lines are essential for computing accurately the structure of model atmospheres and for spectrophotometric predictions (e.g. Short & Lester 1996). But as the quantum mechanics predictions are accurate to only a few per cent, wavelengths for these lines may be largely uncertain. The predicted lines are, therefore, unsuitable for high resolution analyses (Bell, Paltoglou & Tripicco 1994; Castelli & Kurucz 2004; Munari, Sordo, Castelli & Zwitter 2005).

In the last decades Kurucz has made an invaluable contribution, estimating in detail tens of millions of atomic and molecular lines (Kurucz 2011). Many others have tried to improve these lists for the purpose of chemical analysis (Fuhr & Wiese 2006; Meléndez & Barbuy 2009; Den Hartog, Lawler, Sobeck, Sneden & Cowan 2011; Wood, Lawler, Sneden & Cowan 2013). Nevertheless, these approaches only improve the quality of a selective group lines, although it has been shown in the literature (Martins & Coelho 2007) that even manual calibrations of some few specific lines can produce significant improvement on the synthetic spectra generated. While these studies can provide accurate and precise determinations for thousands of lines, tens of millions of lines are estimated to be needed to compute a complete stellar grid with a good range of atmospheric parameters. Therefore, innumerous amount of lines remain poorly characterized.

Aiming to overcome these limitations and to create a more robust and automatic method for the callibration of atomic parameters, we introduce a powerful statistical technique: the cross-entropy algorithm (hereafter, CE). In this work we validate our CE method to calibrate atomic line lists using synthetic spectra of very well known stars. The code for calibrating the atomic line lists using the CE method is called ALiCCE (Atomic Lines Calibration using the Cross-Enthropy algorithm). We show the great capability of the method to determine the atomic parameters of the absorption lines only by comparing the models with the desired spectra. This paper is structured as follows: in Section 2 we explain the Cross-Entropy method, in Section 3 we explain the methodology used in the validation process, in Section 4 we show our results and in Section 5 we present our discussion and conclusion.

2. The Cross-Entropy method and ALiCCE

To compute a synthetic stellar spectrum, the spectral synthesis code reads a model atmosphere, and a list of molecular and atomic transitions. We choose the library from (Castelli & Kurucz 2004) which is based on ATLAS9[1] Kurucz (1970); Sbordone,

[1] Available at
http://wwwuser.oat.ts.astro.it/castelli/sources/atlas9codes.html

Bonifacio, Castelli & Kurucz (2004). To generate the synthetic spectra we chose the code SYNTHE (Kurucz & Avrett 1981) in its public Linus port by Sbordone et al. (2004).

The parameters needed for the computation of each line profile are: the central wavelength, the transition levels energy, the oscillator strength $log(gf)$ (which is related to the line intensity) and the broadening parameters (natural - $log\ \Gamma_{rad}$, Stark - $log\ \Gamma_{StarK}$ and Van der Waals - $log\Gamma_{vdW}$ broadening). In the present work we choose to calibrate the three most uncertain parameters of the list: the oscillator strength and the two pressure broadenings - Stark and Van der Waals.

Although the same line list is used to generate the spectrum of every type of star, not every line of this list will be present in a given stellar spectrum. Ideally, in order to be able to calibrate all the lines in a line list one would have to attempt to calibrate the lists to a variety of stellar spectral types simultaneously.

2.1 The Cross-Entropy algorithm

The CE analysis was originally used in the optimization of complex computer simulation models involving rare event simulations Rubinstein (1997), where very small probabilities have to be accurately estimated. It was modified by Rubinstein (1999) to deal with continuous multi-extremal and discrete combinatorial optimization problems.

The CE method was already used successfully in astrophysical problems like the study of sources with jet precession (Caproni, Monteiro & Abraham 2009), fitting the color-magnitude diagram of open clusters (Monteiro, Dias & Caetano 2010; Oliveira, Monteiro, Dias & Caetano 2013) and modeling very long baseline images (Caproni, Monteiro, Abraham, Teixeira & Toffoli 2011).

The basic procedures involved in the CE optimization can be summarised as follows (e.g. Kroese, Porotsky & Rubinstein 2006):

(i) Random generation of an initial parameter sample, obeying predefined criteria;

(ii) Selection of the best samples based on some mathematical criterion;

(iii) Random generation of updated parameter samples from the previous best candidates to be evaluated in the next iteration;

(iv) Optimization process that repeats steps (ii) and (iii) until a pre-specified stopping criterion is fulfilled.

The relevant equations can be found at Caproni et al. (2011).

2.2 The performance function

In order to select the best model that represents our observation we need to define a performance function, based on the desired characteristics of the solution. In this work the performance function has to be defined as a way to compare model with observed spectra pixel by pixel. To do this comparison we choose as a performance function the combination of the sum of the squared residuals and their respective variance, as suggested in Caproni et al. (2011).

Let the quadratic residual $R_m(k)$ at a given wavelength pixel m and iteration k be defined as the squared difference between the observed spectrum we want to reproduce, I_m, and the generated model spectrum $M_m(k)$ at a k-iteration, i.e. $R_m(k) = [I_m - M_m(k)]^2$. The mean square residual value of the model fitting $\bar{R}(k)$ can be calculated from:

$$\bar{R}(k) = \frac{1}{N_{\rm pixel}} \left[\sum_{m=1}^{N_{\rm pixel}} R_m(k) \right]. \tag{1}$$

As mentioned in the previous section, we need the spectra of stars spanning a range of $T_{\rm eff}$, $\log g$ and metallicity in order to guarantee all lines to be calibrated be present in at least one of them. Considering $N_{\rm s}$ as the number of stars used in the calibration, our tentative model spectra obtained from the $N_{\rm s}$ parameters $\mathbf{x}_i(k)$ at iteration k are ranked through the performance function:

$$S_{\rm prod}(\mathbf{x}_i, k) = \prod_{j=1}^{N_S} \bar{R}(k) \times \frac{1}{N_{\rm pixel}} \left[\sum_{m=1}^{N_{\rm pixel}} \left(R_m(k) - \bar{R}(k) \right)^2 \right], \tag{2}$$

which corresponds to the product of individual merit functions associated with the $N_{\rm s}$ stars used in the calibration. The product of these functions is important to ensure that high luminosity stars do not become more important than low luminosity stars. Note also that, if I_m and M_m are expressed in terms of ergs cm^{2}Å^{-1}, $S_{\rm prod}$ has units of (ergs cm^{2}Å^{-1})^{6N_S}.

2.3 Determination of the parameters and their uncertainties

As commented previously, CE optimization generates N random tentative solutions at each iteration k, selecting the best $N_{\rm elite}$ set of model parameters in terms of the values of $S_{\rm prod}$.

If the stopping criteria is related to a maximum number of iterations, the last iteration might not be exactly the one with the best $S_{\rm prod}$. The best values of the parameters p_i^*, as well as their respective uncertainties $\sigma_{p_i^*}$, were then determined as

follows:

$$p_i^* = \sum_{k=1}^{k_{\rm conv}} w_k p_{ik} \left(\sum_{k=1}^{k_{\rm conv}} w_k \right)^{-1} \tag{3}$$

and

$$\sigma^2_{p_i^*} = \sum_{k=1}^{k_{\rm conv}} w_k \, (p_{ik} - p_i^*)^2 \left(\sum_{k=1}^{k_{\rm conv}} w_k \right)^{-1} , \tag{4}$$

where $k_{\rm conv}$ is the number of iterations where the parameters converged, and p_{ik} represents the set of model parameters that produce the minimum value of $S_{\rm prod_k}$ among all tentative solutions at iteration k and $w_k = S^{-2}_{\rm prod_k}$. The power index -2 in the definition of w_k was adopted in order to make the tentative solutions with the lowest values of $S_{\rm prod_k}$ more important in the calculation of p_i^*. The value of $k_{\rm conv}$ corresponds to the iteration where the value of p_{ik} did not change after a certain number of iterations. We found that a safe value for this number is 10% of the total number of iterations.

3. Validating ALiCCE

For this initial validation step we choose the range from 851.0 to 853.0 nm, which is a region relatively free of strong molecular transitions. For this range of only 20 Å there are almost 290 atomic lines that can be attempted for calibration. It is unfeasible to calibrate all these lines manually for hundreds or thousands of angstroms as a stellar library would need.

To validate the calibration technique we used synthetic spectra instead of real ones. These spectra were generated with the original line list of Kurucz, so, in this way, we know the values of the atomic parameters we are looking for. To make the test more realistic we also introduced some random noise in these spectra. The S/N we used here is 400 and is smaller than what we actually expect for the real observations that will be used for the calibrations. The idea is to verify if ALiCCE is able to recover the atomic parameters without any knowledge about their values, only through the comparison of the spectra.

The stars that will be used in the calibration have a crucial role. As mentioned, the ideal case would be to use many stars of different spectral types. However, in practice, this is unfeasible. The stellar spectra that will be used for the calibration have to be from stars with very well determined atmospheric parameters not depending upon model. The spectra will also have to have very high resolution and signal-to-noise. In practice, there are very few available spectra in the literature that satisfy these criteria. Here we generated the synthetic spectra with the atomic parameters of real stars that will be used in the actual calibration: three well known stars, that have not only very well determined atmospheric parameters but also have very different spectral types: The Sun (G2V), Acturus (K1.5III) and Vega (A0V).

Table 1. Participation of some lines in the spectra generated by SYNTHE.

λ (nm)	Sun Index	Vega Index	Arcuturs Index	Sum
851.3916	0.9936	0.9998	1.0000	2.9934
851.4005	0.9993	0.9978	0.9975	2.9946
851.4072	0.1365	0.3726	0.9931	1.5022
851.4599	0.8285	0.7805	0.9972	2.6062
851.4629	0.9998	1.0000	1.0000	2.9998

3.1 Methodology and definition of CE parameters

The performance function was defined in a way that it compares models (spectra generated by ALiCCE) with observations (input synthetic spectra with noise, for the validation test), pixel by pixel. The number of tried solutions (N), the number of maximum iterations (k_{max}), and consequently the computation time, are directly related with the number of free parameters to be adjusted. This means that if we choose a large wavelength range to perform the calibration at once the problem might become unfeasible. The calibration can be performed in small steps of wavelengths at a time. The only compromise that must be respected is to choose a wavelength range which is large enough to encompass the entire lines that will be calibrated. For this calibration process we performed the calibration in steps of 1.5 Å width, with an overlap of 0.5 Å width from one to another.

Another constrain that has to be defined is the interval in which each atomic parameter can vary in the first iteration. Errors in the atomic parameters can be as high as 400 %, so that was the initial interval chosen for them. For the remaining CE parameters, after many tests we adopted k_{max} = 2000, α = 0.5, q = 6, N=1000 and N_{elite} = 0.05N for the calibration of each 1.5 Å, which were the most efficient for the performance of the technique.

4. Results

By the end of the calibration process we expect to recover the atomic parameters used to generate the synthetic spectrum with noise for the lines that are present with a reasonable strength in the stars used in the calibration. If a line is too weak or not present in at least one of the stars, we don't expect to recover its original parameters. Since we used synthetic spectra we have prior information about which lines are expected to be recovered by our code. SYNTHE generates an output in which it tabulates the participation of each line in the spectra. The numbers correspond to the per mil residual intensity at line center if the line were computed in isolation. Table 2 shows examples of these values for some of the lines in our interval, for a sensibility of 10^{-4}. The closer the indexes are to 1.0, the weaker is the line.

Amongst the parameters the code has to calibrate, $log(gf)$ was the one with higher success of recovery. This is not surprising as $log(gf)$ usually is the main parameter that drives the line profile. For the broadening parameters the recovery rate was much lower. The recovery of the broadening parameters is extremely dependent on the stellar rotational velocity. If the rotational velocity is too high, these parameters cannot be recovered, since the Doppler broadening effect will dominate the line profile. For the rotational velocity of the stars used in these calibrations we can recover the broadening parameter only for very strong lines.

As an example of the convergence behavior, Fig. 1 shows the $log(gf)$ recovery for the lines listed in Table 1. In this figure, the blue dotted line shows the expected value of $log(gf)$, and the black line shows the evolution of the values found by ALiCCE as a function of iteration. The only two lines that converged to the expected value in this figure are lines 851.4072 and 851.4599. However, line 851.3916 seems to have converged to a value which is not the nominal value for this line. In this validation test we have this information, but when applying the code for real stars, how will it be possible to distinguish between the corrected values and the incorrect ones? It is clear that a simple visual inspection of Fig. 1 is not enough.

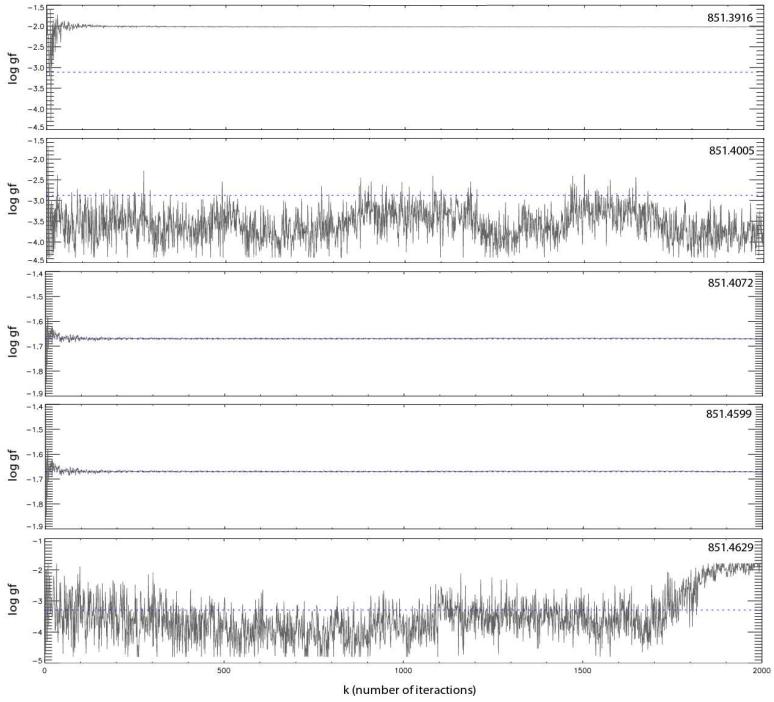

Figure 1. Example of the evolution of the values of log gf found by ALiCCE at each iteration (black line) for the lines in table 1. The doted blue line shows the expected value.

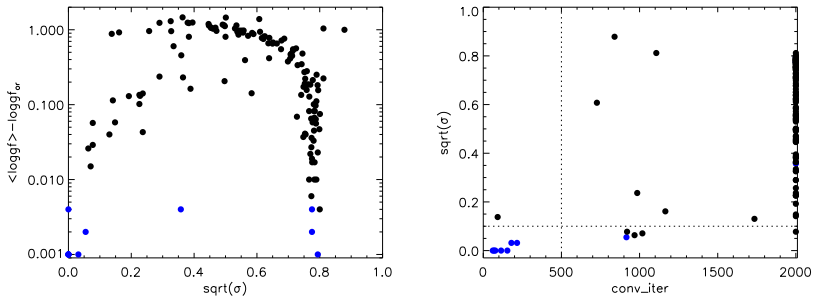

Figure 2. Convergence criteria. The top panel shows the difference between expected values and the values found by ALiCCE for each line in the interval, as a function of the errors calculated for the calibration process. The bottom panel shows the errors from the calibration process as a function of the iteration in which the line converged to a single value. Lines that converged to the correct value are located in the bottom left of this figure.

The criteria to select between the correct converged values and the incorrect ones are shown in Fig. 2. The left figure shows the difference between the average value p_i^* found from the synthesis and the expected value as a function of their uncertainties $\sigma_{p_i^*}$. From this figure it becomes clear that parameters that converged to their expected values have small uncertainties. The right figure shows this effect more clearly: lines that converged to their expected values do so in the first iterations. Using this as a criteria, we can identify lines that actually converged to an expected value even if we don't know this value a priori, or have any knowledge of their participation in the spectra of the stars. Given these results we believe that ALiCCE is an efficient tool for the calibration of atomic line lists for stellar spectrum synthesis.

5. Conclusion

In this work we present the validation of a method for calibrating atomic line lists used by spectral synthesis codes to generate stellar spectra called ALiCCE, which uses the Cross-Entropy algorithm. For validation of our technique we simulated a calibration against the synthetic spectra with noise of three well known stars with very well determined atmospheric parameters: the Sun, Arcturus and Vega.

The results obtained show that the Cross-Entropy algorithm is effective to recover the atomic lines parameters used in line lists of spectral synthesis codes for lines that contribute at least with of the continuum flux. The code was able to recover the $log(gf)$ of the lines that had reasonable signal in at least one of the three stars. In addition, the code is able to recover broadening parameters of very strong lines, when they are present.

Acknowledgements

This research has been partially supported by the Brazilian agency FAPESP (2011/00171-4 and 2013/14421-8). L.M. also thanks CNPq through grant 305291/2012-2.

References

Bell R. A., Paltoglou G., Tripicco M. J., 1994, MNRAS, 268, 771
Caproni A., Monteiro H., Abraham Z., 2009, MNRAS, 399, 1415
Caproni A., Monteiro H., Abraham Z., Teixeira D. M., Toffoli R. T., 2011, ApJ, 736, 68
Castelli F., Kurucz R. L., 2004, A&A, 419, 725
Den Hartog E. A., Lawler J. E., Sobeck J. S., Sneden C., Cowan J. J., 2011, ApJS, 194, 35
Fuhr J. R., Wiese W. L., 2006, J. Phys. Chem. Ref. Data., 35, 1669
Kroese D. P., Porotsky S., Rubinstein R. Y., 2006, Methodol. Comput. Appl. Probab., 8, 383
Kurucz R. L., 1970, SAO Special Report, 309
Kurucz R. L., 1992, RMxAA, 23, 45
Kurucz R. L., 2011, Canadian Journal of Physics, 89, 417
Kurucz R. L., Avrett E. H., 1981, SAO Special Report, 391
Martins L. P., Coelho P., 2007, MNRAS, 381, 1329
Meléndez J., Barbuy B., 2009, A&A, 497, 611
Monteiro H., Dias W. S., Caetano T. C., 2010, A&A, 516, A2
Munari U., Sordo R., Castelli F., Zwitter T., 2005, A&A, 442, 1127
Oliveira A. F., Monteiro H., Dias W. S., Caetano T. C., 2013, A&A, 557, A14
Rubinstein R. Y., 1997, European Journal of Operational Research, 99, 89
Rubinstein R. Y., 1999, Methodology and Computing in Applied Probability, 2, 127
Sbordone L., Bonifacio P., Castelli F., Kurucz R. L., 2004, Memorie della Societa Astronomica Italiana Supplementi, 5, 93
Short C. I., Lester J. B., 1996, ApJ, 469, 898
Wood M. P., Lawler J. E., Sneden C., Cowan J. J., 2013, ApJS, 208, 27