

Recovering UV atomic data with evolutionary algorithms

From noise-free feasibility tests to realistic conditions

Joshua E. Lara-Sabala, Emanuele Bertone, Alicia Morales-Reyes, Miguel Chávez

INAOE, Puebla, Mexico



THE PROBLEM

UV atomic data limits accurate spectral synthesis

$10^2 - 10^3$ lines / Å

in the near UV, so line profiles overlap and blend.

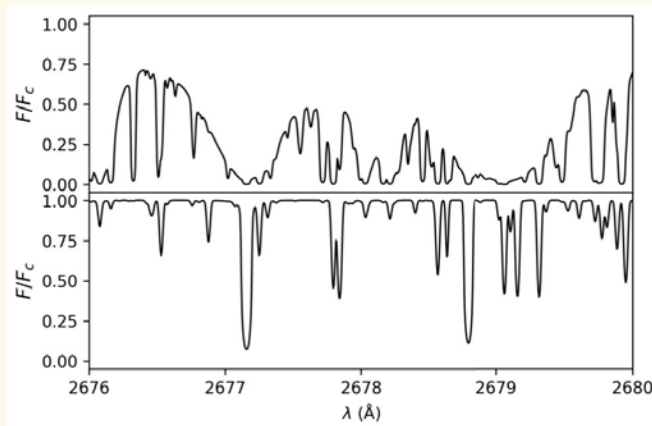
$\geq 25\%$

uncertainty for most f-values (Wahlgren 2011).

> 2 dex

log gf differences between databases
(Laverick et al. 2018).

Automated efforts to recover astrophysical atomic parameters have been quite limited so far, and only in low-blending regions: Martins et al. 2014 (CE, 8510–8530 Å), Shetrone et al. 2015 (1.5–1.7 μm), Laverick et al. 2019 (4200–6800 Å). None in the UV.

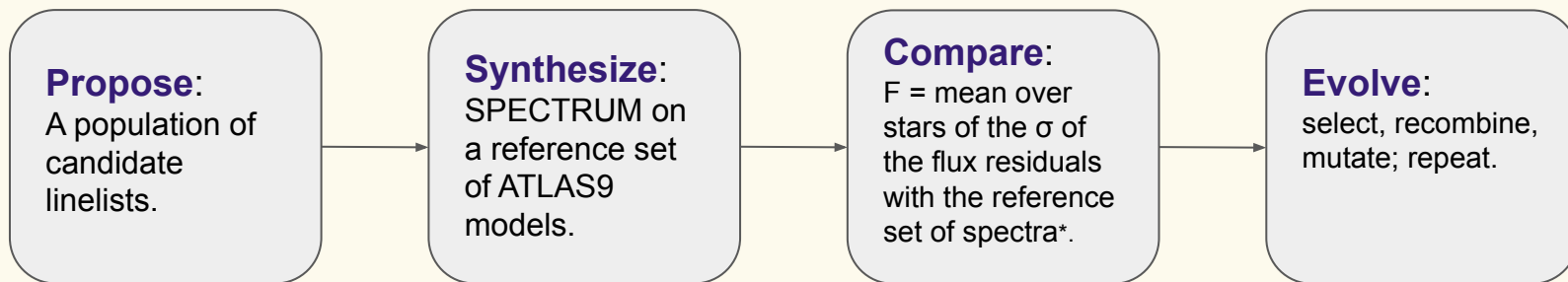


Synthetic spectra: $T_{\text{eff}} = 5000\text{ K}$ (top) vs $10\,000\text{ K}$ (bottom).

METHOD

Fit every line in a window at once with Computational Intelligence algorithms

Each candidate solution is one vector that contains four parameters per line: λ_c (central wavelength), $\log g_f$ (oscillator strength), $\log \Gamma_{\text{Stark}}$ and $\log \Gamma_{\text{VdW}}$ (pressure broadening parameters). The search space is $\pm 3\sigma$ around VALD, with σ taken from a VALD–Kurucz cross-match.



* This is a feasibility test: the “observations” are synthetic spectra from a perturbed linelist (LRP), so the true answer is known.

SETUP

A controlled test for algorithm performance comparison

10 Å

λ 2700–2710 Å, plus 1.5 Å
for strong-line wings.

5 stars*

Sun, α Cen A, Procyon,
Vega, Sirius A (G–A MS).

500 000

resolving power,
noise-free spectra.

1273

free parameters with
impact $I \geq 0.005$ ($\approx$ HST
noise).

180 000

objective-function
evaluations per run.

5 × 5

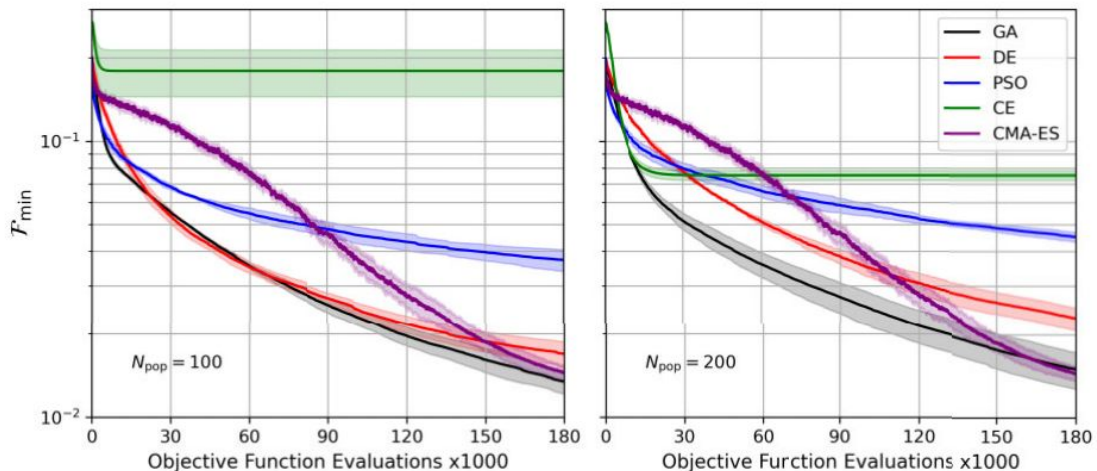
algorithms × independent
runs.

* Synthetic spectra computed with the perturbed linelist (LRP), so the true parameters are known.

ALGORITHMS

GA and CMA-ES lead

Lara-Sabala et al. 2026



Best F vs evaluations; $N_{pop} = 100$ (left), 200 (right), for CMA-ES the population size was defined with an empirical formula; band = σ over 5 runs.

GA

$$F = 1.34 \times 10^{-2} \pm 0.13 \times 10^{-2}$$

CMA-ES

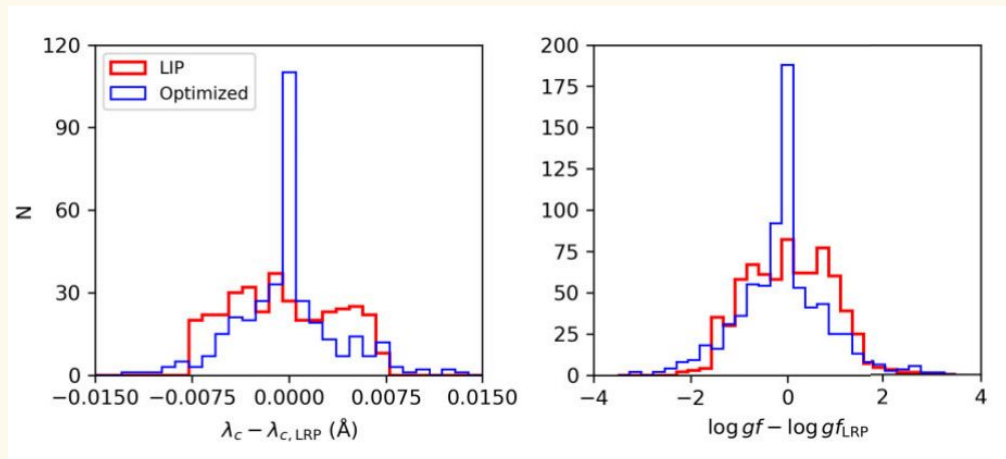
$$F = 1.45 \times 10^{-2} \pm 0.09 \times 10^{-2}$$

ANOVA: $p = 0.22$, statistically equivalent

CE stalls early; PSO and DE are slower. GA and CMA-ES are still improving at 180k evaluations.

ACCURACY

High-impact lines are recovered tightly



Offset from the true value: start (LIP, red) vs GA (blue)

σ of offset	λ_c (Å)	log gf
Start, all	0.0041	0.851
GA, all	0.0038	0.882
GA, $I > 0.4$	0.0016	0.665
GA, $I > 0.8$	0.00022	0.531

Medians near zero: 1×10^{-4} Å and -0.045 dex. The gain concentrates where lines actually shape the spectrum.

RECOVERY

Recovered with accuracy and precision

Recovered = the true value lies within $1\sigma_{\text{run}}$ of the 5 run result, and $\sigma_{\text{run}} \leq$ the σ used to build the reference.

63%

of λ_c for lines with $I > 0.4$
(30 % over all 355)

Against the databases

λ_c uncertainty $\leq$ NIST: **19 / 28**

log gf class $\geq$ NIST: **26 / 36**

log gf uncertainty $\leq$ VALD: **12 / 21**

48%

of log gf for lines with $I > 0.4$
(25 % over all 741)

Damping constants: only 5 / 79 Stark and 8 / 98 Van der Waals recovered, owing to blending and log gf - Γ degeneracy in saturated lines.

STEP 2. MORE REALISTIC SCENARIO

Testing sources of uncertainty

Same pipeline using the GA.

1. Noise and Resolution

Coupled, tested together on scenarios based on real instruments.

2. Atmospheric parameters

Errors in T_{eff} , $\log g$ and $[M/H]$, and a larger set of reference stars.

3. Chromosphere

Line cores of strong lines can carry chromospheric emission that photospheric LTE models cannot reproduce.

4. Continuum placement

Systematic normalization bias, typical of the crowded UV.

5. Unidentified transitions

Real spectra hold lines missing from the linelist; the GA may compensate by distorting nearby known lines.

BLOCK 1

Noise and resolution grid

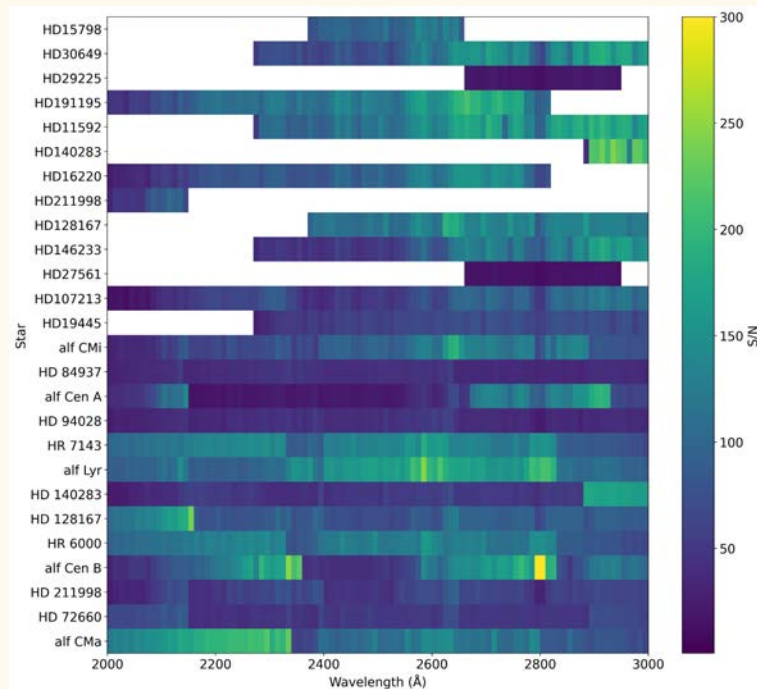
Resolution: E230H ($R \approx 114\,000$) and E230M ($R \approx 30\,000$), the two HST/STIS near-UV echelle modes.

Noise: $S/N = 50, 100, 150$ and 200 , with 5 independent noise realizations per case.

For this we convolve to the instrument profile, then add Gaussian noise.

Output: recovery fraction vs line strength over ($R, S/N$).

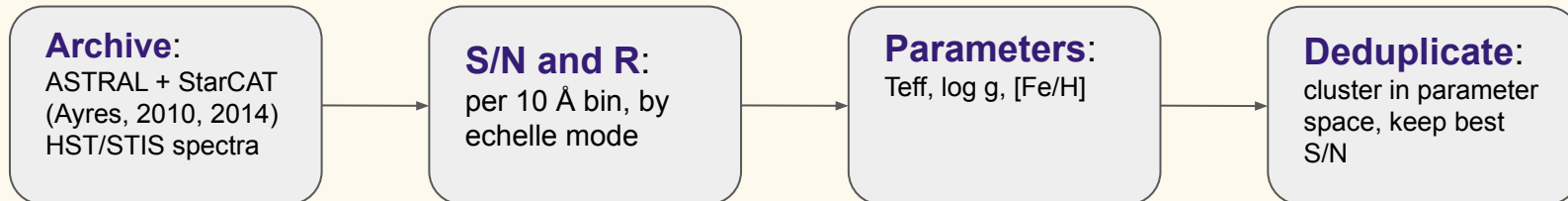
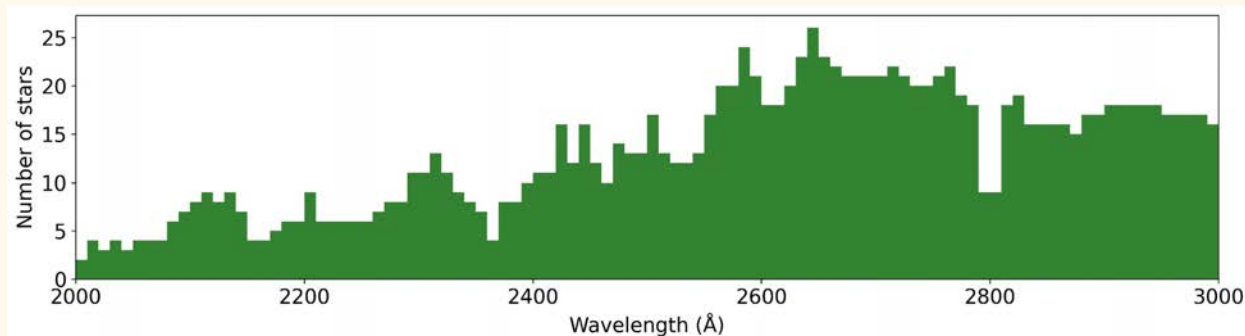
Goal: an empirical R and S/N threshold to construct the final reference set of observations.



S/N per 10 Å bin across 2000–3000 Å for some of the candidate reference stars, estimated from the 98–100th flux percentile as a continuum proxy (Ayres, 2010, 2014).

BLOCK 2

Building the reference star sample



Number of stars with $S/N > 100$ per 10 Å bin

CONCLUSIONS

Takeaways and next steps

So far

- Simultaneous EA fitting of UV line parameters is feasible; $GA \approx$ CMA-ES.
- High-impact lines: 63 % of λ_c and 48 % of $\log gf$ recovered, often at or below NIST/VALD uncertainties.
- Damping constants stay degenerate with $\log gf$ and blends.

Next

- Establish final star sample.
- Assessment of different sources of uncertainties.
- Application to real observations and construction of a new UV linelist.
- Building a new stellar spectral library in the UV.

SOURCES

References

Ayres T. R., 2010, ApJS, 187, 149 (StarCAT)

Ayres T. R., 2014, ASI Conf. Ser., 11, 1 (ASTRAL)

Gray R. O., 1999, ASCL, ascl:9910.002 (SPECTRUM)

Kramida A. et al., NIST Atomic Spectra Database

Kurucz R., 1993, ATLAS9, Kurucz CD-ROM No. 13,
SAO

Lara-Sabala J. E. et al., 2026, MNRAS, 548, 1

Laverick M. et al., 2018, A&A, 612, A60

Laverick M. et al., 2019, Proc. IAU Symp. 350, 386

Martins L. P. et al., 2014, MNRAS, 442, 1294

Ryabchikova T., Pakhomov Y., Piskunov N., 2018,
Galaxies, 6, 93 (VALD)

Shetrone M. et al., 2015, ApJS, 221, 24

Soubiran C. et al., 2016, A&A, 591, A118 (PASTEL)

Wahlgren G., 2011, Can. J. Phys., 89, 345