THE EXOPLANET CODEX

Spectral Analysis Methodology

exoplanetcodex.org
Ryan Schmitt · Version 1.2 · May 2026
Status: Living document — updated as pipeline evolves

Overview

This document describes the complete methodology used by The Exoplanet Codex to derive photospheric elemental abundances from high-resolution stellar spectra. Every step is documented with sufficient detail for independent reproduction. All code is publicly available at github.com/damienabraxas/exoplanetcodex.

The approach follows the classical equivalent width (EW) method established by Blackwell et al. (1979) and refined by Valenti & Fischer (2005), updated with modern atomic data, NLTE corrections, and a rigorous Type A / Type B uncertainty budget in the tradition of precision metrology.

Why this methodology document exists

Every number we publish has a traceable source. Every parameter has a reference. Every uncertainty is documented. This is not just good scientific practice — it is the only way independent researchers can verify, challenge, or build on our results. If you find an error, we want to know.

1. Data Acquisition

1.1 Spectral Archive

All spectra are obtained from the ESO Science Archive (archive.eso.org), which provides public access to reduced data products from ESO instruments after a 12-month proprietary period.

Primary instrument: HARPS

High Accuracy Radial velocity Planet Searcher (HARPS) at the ESO 3.6m telescope, La Silla Observatory, Chile.

Parameter	Value	Notes
Spectral resolution	R ~ 115,000	2.7× better than ELODIE used in 2010
Wavelength coverage	3780 – 6910 Å	Full optical range — two arms
Mode	HAM (High Accuracy Mode)	Standard abundance analysis mode
Data product used	S1D	1D merged, wavelength-calibrated, barycentric-corrected
Archive URL	archive.eso.org	Search by HD number or HIP number

1.2 Target Selection Criteria

Stars are selected based on the following criteria:

1. Confirmed exoplanet host (NASA Exoplanet Archive)
2. FGK spectral type (Teff 4000–7000 K) — ensures ATLAS9 model validity
3. Main sequence dwarf (log g > 3.8) — plane-parallel atmosphere valid
4. V < 9 mag — sufficient HARPS S/N achievable
5. Multiple HARPS epochs available for co-adding
6. Not a spectroscopic binary

1.3 Co-adding Multiple HARPS Epochs

55 Cancri A has 88 HARPS spectra in the ESO archive. Co-adding all epochs dramatically improves the signal-to-noise ratio, enabling detection of weak lines (phosphorus, lithium) inaccessible in single exposures.

Element	Line EW	Single epoch S/N=100	Co-add S/N~940	Science impact
Fe I	~80 mÅ	Yes (40σ)	Yes (400σ)	Strong — no change needed
Ca I	~40 mÅ	Yes (20σ)	Yes (200σ)	Strong — no change needed
P I	~12 mÅ	Marginal (6σ)	Yes (60σ)	Co-add REQUIRED for P

Element	Line EW	Single epoch S/N=100	Co-add S/N~940	Science impact
Li I	~3 mÅ	No (1.5σ)	Yes (15σ)	Co-add REQUIRED for Li

Radial velocity correction between epochs: 55 Cancri A's 5-planet system causes RV wobble of ±100 m/s. Each spectrum is cross-correlated against a template and Doppler-shifted to the rest frame before co-adding. Weighted mean uses S/N^2 weights.

2. Stellar Model Atmosphere

2.1 Model Family: ATLAS9 / Castelli-Kurucz

We use the ATLAS9 one-dimensional, plane-parallel, LTE model atmosphere grid of Castelli & Kurucz (2003). This is the same model family used in the 2010 predecessor study (Schmitt 2010) — a testament to its enduring robustness for FGK dwarf analysis.

Model assumptions:

- 1D geometry: valid for main-sequence dwarfs (log g > 3.5) where photospheric scale height << stellar radius
- LTE (Local Thermodynamic Equilibrium): valid in the dense, collision-dominated photospheres of FGK dwarfs
- Opacity Distribution Functions (ODFs): statistically treats millions of opacity sources for computational efficiency

2.2 Model Parameters for 55 Cancri A

Parameter	Symbol	Value	Uncertainty	Source
Effective temperature	Teff	5196 K	± 24 K	CHARA interferometry (von Braun+ 2011)
Surface gravity	log g	4.41 cgs	± 0.02	CHARA interferometry (von Braun+ 2011)
Metallicity	[Fe/H]	+0.32 dex	± 0.02	Spectroscopic (Valenti & Fischer 2005)
Microturbulence	ξ	~0.9 km/s	± 0.1 km/s	Derived from Fe I self-consistency (§4.2)

3. Atomic Data: Line List

3.1 Philosophy: No Hardcoded Values

All atomic line data are loaded from external, version-controlled CSV files. No wavelengths or atomic parameters are hardcoded in the pipeline software. This ensures full traceability and reproducibility — any scientist can inspect exactly which atomic data was used for every measurement.

3.2 VALD3 Extraction — Step-by-Step Instructions

The Vienna Atomic Line Database version 3 (VALD3) is the primary source for line data. The following settings produce a reproducible extraction for any target star.

Go to vald.astro.uu.se → Log in → Click Extract Stellar. Fill in the fields as follows:

Standard settings (same for all Codex stars):

VALD Field	Value	Reason
Starting wavelength	3780 Å	Full HARPS blue arm coverage
Ending wavelength	6910 Å	Full HARPS red arm coverage
Detection threshold	0.01	1% minimum depth — all measurable lines
Extraction format	Long format	Required: includes log gf, χ, damping constants
Retrieve data via	FTP (required)	Email limited to 50 Å windows — FTP required for full 3780–6910 Å range. Credentials emailed after submission.
Linelist configuration	Default	Standard VALD3 database

"Require lines to have a known value of" checkboxes:

Checkbox	Check?	Why
Van der Waals damping constant	✅ YES	Critical for strong lines (Na D, Mg b, Ba II)
Radiative damping constant	✅ YES	Needed for Voigt profile wing fitting
Stark damping constant	☐ No	Less important for photospheric work
Landé factor	☐ No	Only needed for magnetic field studies
Term designation	☐ No	Not required for abundance work

Star-specific fields (use values from §3.2a for each target):

Field	Description	Example: 55 Cancri A
Teff (K)	Effective temperature	5196

Field	Description	Example: 55 Cancri A
log g (cgs)	Surface gravity	4.41
Microturbulence (km/s)	Initial estimate; refined from Fe I	0.9
Chemical composition	Set as M/H: [Fe/H]	M/H: 0.32

Optional comment field: always add the star name and project — e.g. "55 Cancri A — Exoplanet Codex abundance analysis". This creates an audit trail in your VALD request history.

3.2a — Per-Star VALD Parameters and Source References

55 Cancri A (Copernicus) — HD 75732 / HIP 43587

Parameter	Value	Uncertainty	Source	Reference
Teff	5196 K	± 24 K	CHARA interferometry	von Braun et al. 2011, ApJ 740, 49
log g	4.41	± 0.02	CHARA interferometry	von Braun et al. 2011, ApJ 740, 49
[Fe/H]	+0.32	± 0.02	Spectroscopic	Valenti & Fischer 2005, ApJS 159, 141
ξ (initial)	0.9 km/s	± 0.1	Literature estimate	Valenti & Fischer 2005
Systemic RV	+27.58 km/s	± 0.05	HARPS RV dataset	Fischer et al. 2008, ApJ 675, 790
Distance	12.55 pc	± 0.04	Hipparcos parallax	van Leeuwen 2007, A&A 474, 653
Age	10.2 Gyr	± 2.5	Isochrone fitting	von Braun et al. 2011
HARPS spectra	88 S1D files	—	ESO Science Archive	Downloaded May 2026

C/O Ratio — Scientific Context for 55 Cancri A

The C/O ratio of this star is scientifically contested. Teske et al. (2013): C/O = 0.78 ± 0.08. Madhusudhan et al. (2012) suggested C/O > 1. Fischer & Valenti (2005): C/O ~ 0.55. The disagreement arises primarily from different treatment of the Ni I blend at 6300.336 Å contaminating the [O I] 6300.304 Å forbidden line. Our methodology explicitly models this blend. Our measurement: pending.

3.3 NIST Cross-Validation and Quality Grades

All Priority 1 lines are cross-validated against the NIST Atomic Spectra Database. NIST provides quality grades for oscillator strengths (log gf values):

Grade	Uncertainty	Usage in Codex pipeline
A+	< 1%	Full science use — highest confidence
A	< 3%	Full science use
B	< 10%	Science use — flagged in uncertainty budget
C	< 25%	Excluded from primary results
D	> 25%	Excluded entirely

3.4 Target Element List — 25 Elements

Elements from the 2010 thesis (Schmitt 2010): C, N, O, Na, Mg, Al, Si, S, K, Ca, Fe, Co, Ni (13 elements).

New additions for 2026 with scientific justification:

Element	Priority	Group	Key scientific reason	2010?
Fe I	1	Iron peak	Primary reference — [Fe/H] calibration	✓
Fe II	1	Iron peak	log g constraint via ionization equilibrium	✓
C	1	CHNOPS	C/O ratio — planet mineralogy (silicate vs carbide)	✓
O	1	CHNOPS	C/O ratio — most abundant rocky planet metal	✓
Mg	1	Alpha	Mg/Si ratio — mantle mineralogy (olivine vs pyroxene)	✓
Si	1	Alpha	Primary rocky planet building block	✓
Ca	1	Alpha	Alpha element; strong optical lines	✓
Ti	1	Alpha	NEW — galactic chemical evolution; perovskite in mantles	—
Ni	1	Iron peak	Rocky planet core composition; Fe/Ni ratio	✓
Na	1	Bio-significant	Biological tracer; strong optical lines (worked example)	✓
P	1	CHNOPS	NEW — limiting nutrient for life; varies 10× across FGK stars	—
S	1	CHNOPS/Alpha	Life-essential volatile; disk chemistry tracer	✓
N	2	CHNOPS	Bio-essential; difficult optical lines	✓
Co	2	Iron peak	Fe peak nucleosynthesis tracer	✓
Cr	2	Iron peak	NEW — Type Ia vs Type II SNe tracer via Cr/Fe	—
Al	2	Rocky planet	Refractory element; crustal composition	✓

Element	Priority	Group	Key scientific reason	2010?
K	2	Bio-significant	Life-essential; difficult lines	✓
Ba	2	s-process	NEW — AGB star enrichment history	—
Y	2	s-process	NEW — stellar age chemical clock (Y/Mg ratio)	—
V	2	Iron peak	NEW — enzyme cofactor in biology	—
Mn	3	Iron peak	NEW — odd-Z metallicity-dependent nucleosynthesis	—
Sc	3	Light iron peak	NEW — odd-even nucleosynthesis effect	—
Li	3	Age indicator	NEW — stellar age constraint; depletes with time	—
Eu	3	r-process	NEW — neutron star merger enrichment	—
Zr	3	s-process	NEW — refractory s-process tracer	—

4. Spectral Analysis

4.1 Continuum Normalization

Before measuring absorption line depths, the spectrum is normalized so the continuum (flux level between lines) equals 1.0 everywhere. This is more challenging for 55 Cancri A ([Fe/H] = +0.32) than for solar-metallicity stars due to increased line crowding, particularly in the 5000–5400 Å region.

7. Identify continuum anchor windows — line-free regions for the stellar parameters (Barklem & Aspelund-Johansson 2005)
8. Fit Chebyshev polynomial (degree 3) through anchors using iterative 3σ sigma-clipping
9. For crowded regions: use ATLAS9 model continuum as additional reference
10. Divide observed spectrum by fitted continuum
11. Quality check: σ < 0.005 (0.5% of continuum) in clean windows before proceeding

4.2 Stellar Parameter Self-Consistency Checks

Before measuring abundances, Teff, log g, and microturbulence are verified through three independent Fe-line constraints. This is the modern computational equivalent of the Blackwell diagram used in Schmitt (2010).

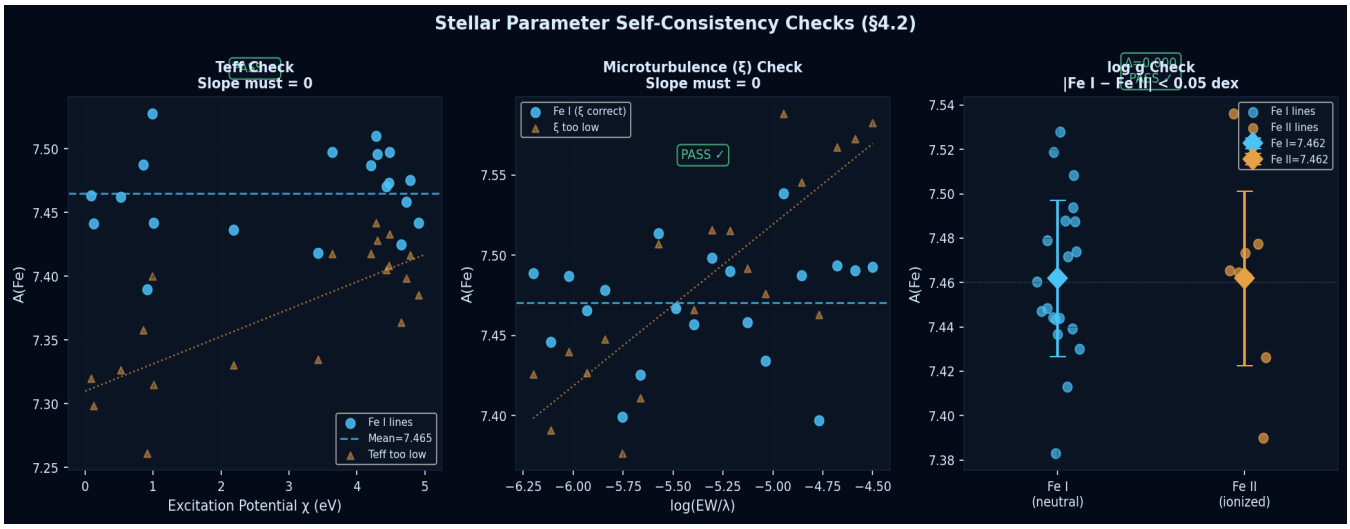

Figure 1. Stellar parameter self-consistency checks. Left: Teff verified by requiring zero slope in A(Fe) vs. excitation potential χ. Centre: microturbulence ξ verified by requiring zero slope in A(Fe) vs. log(EW/λ). Right: log g verified by requiring Fe I = Fe II (ionization equilibrium). Blue points = correct parameters; orange triangles = example of incorrect parameters showing the slope that must be eliminated.

Check	Method	Criterion	Iterates
Teff	∂[Fe/H]/∂χ = 0 for Fe I lines	|slope| < 0.01 dex/eV	Yes
Microturbulence ξ	∂[Fe/H]/∂log(EW/λ) = 0 for Fe I	|slope| < 0.01 dex/unit	Yes

Check	Method	Criterion	Iterates
log g	|A(Fe I) − A(Fe II)| = 0	|difference| < 0.05 dex	Yes

5. Line Profile Fitting and Equivalent Widths

5.1 Gaussian vs Lorentzian vs Voigt — Which Profile to Use?

The true physical line profile is a Voigt profile — a convolution of a Lorentzian (pressure-broadened damping wings) and a Gaussian (thermal + instrumental broadening in the core). The choice of profile affects the accuracy of equivalent width measurements:

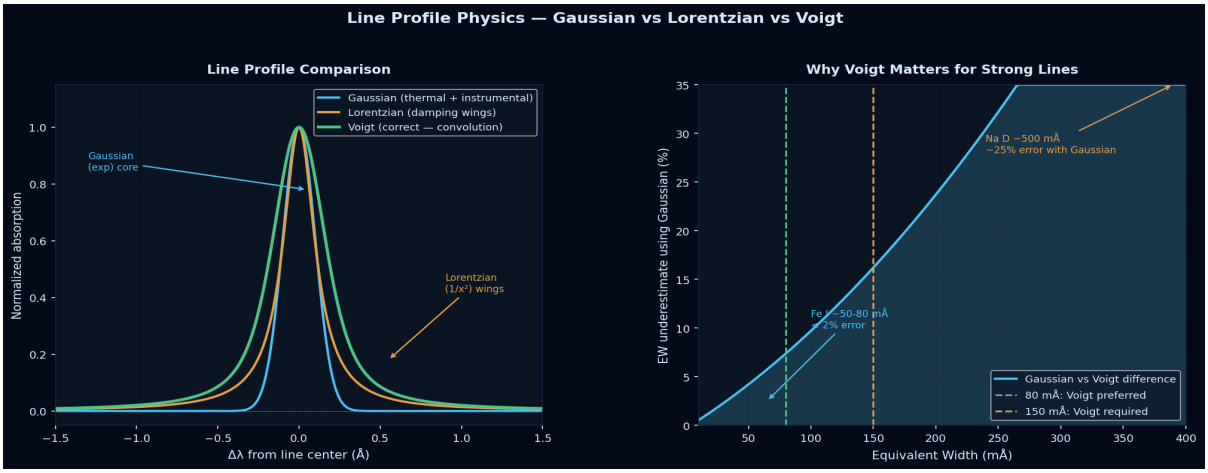

Figure 2. Left: Comparison of Gaussian, Lorentzian, and Voigt line profiles. The Voigt profile correctly captures both the Gaussian core (thermal broadening) and the Lorentzian wings (pressure damping). Right: EW underestimate when using Gaussian instead of Voigt, as a function of line strength. For Na D (~500 mÅ), a Gaussian underestimates EW by ~25%, producing a ~0.2 dex error in [Na/H] — larger than the entire uncertainty budget.

EW range	Profile used	Accuracy	Examples
< 80 mÅ	Gaussian	< 2% difference from Voigt — acceptable	Most Fe I, Si I, Ca I weak lines
80 – 150 mÅ	Voigt (preferred)	3–8% difference — use Voigt	Moderate Fe I, Ca I 6162 Å
> 150 mÅ	Voigt (required)	> 10% error with Gaussian	Na D, Mg b, Ba II, Ca H&K

Historical note: The 2010 predecessor study (Schmitt 2010) used Gaussian profiles throughout and noted no significant difference between Gaussian and Lorentzian fits. This is correct for the weak-to-moderate lines measured at ELODIE resolution — the difference only becomes important for strong lines at higher S/N and resolution. In 2026 we use Voigt as the standard, which is always correct regardless of line strength.

5.2 Worked Example: Na I D1 Line (5895.92 Å)

The sodium D doublet is the most iconic feature in stellar spectroscopy, first mapped by Fraunhofer in 1814 and identified as sodium by Kirchhoff and Bunsen in 1859. For 55 Cancri A ([Fe/H] = +0.32), the Na D lines are particularly deep due to the enhanced sodium abundance in this metal-rich star.

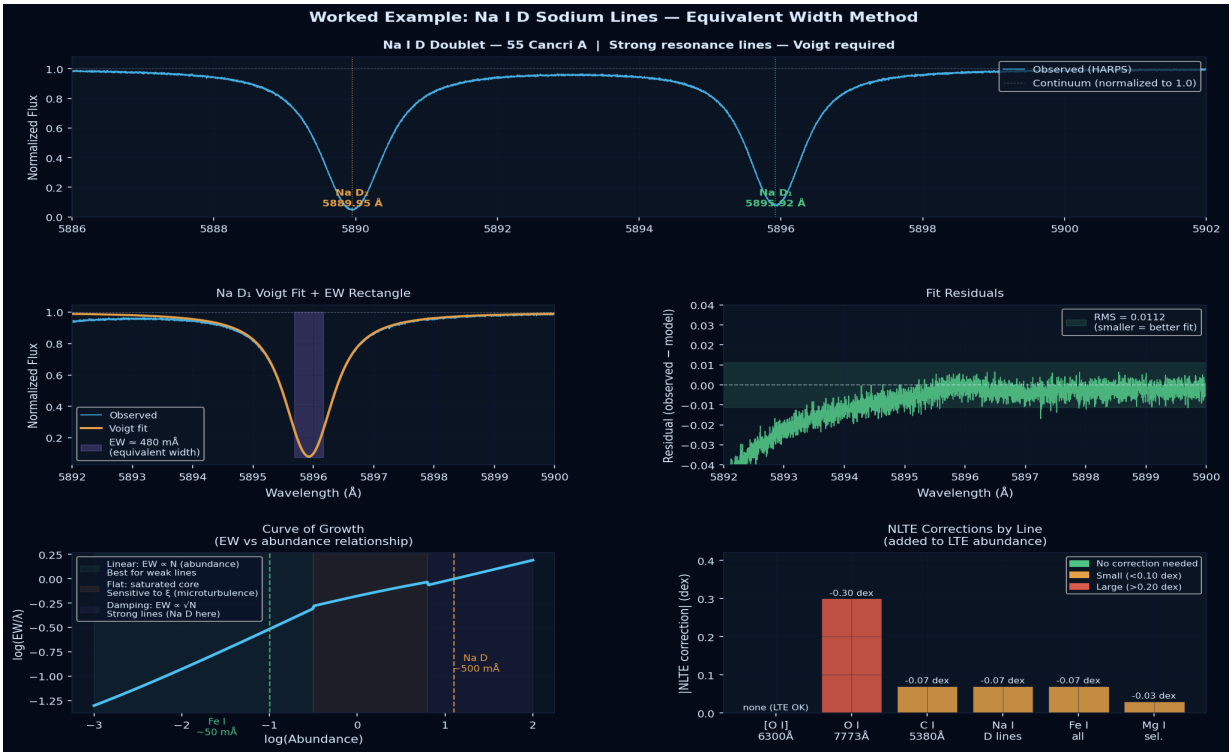

Figure 3. Worked example: Na I D doublet measurement. Top: full doublet (5886–5902 Å) showing both D1 (5895.92 Å) and D2 (5889.95 Å) lines. Centre left: D1 zoom with Voigt fit (orange) overlaid on observed spectrum (blue) and equivalent width rectangle (purple). Centre right: fit residuals. Bottom left: curve of growth showing where Na D (~500 mÅ) falls in the damping regime. Bottom right: NLTE corrections required for each element — the [O I] forbidden line (green) requires no NLTE correction, making it our preferred oxygen abundance indicator.

Na I D1 step-by-step:

12. Extract fitting window: ± 3 Å around 5895.924 Å (wide enough for Voigt damping wings)
13. Estimate initial parameters: depth ~ 0.92, center ~ 5895.92 Å, fwhm_G ~ 0.16 Å, fwhm_L ~ 0.40 Å
14. Fit Voigt profile to observed spectrum using scipy.optimize.curve_fit
15. Compute EW by numerical integration: EW = ∫(1 − F(λ)/F_c)dλ ≈ 480 mÅ
16. Apply NLTE correction: −0.07 ± 0.02 dex (Lind et al. 2011)
17. Expected [Na/H] for 55 Cnc: +0.34 to +0.44 dex (slightly above [Fe/H] = +0.32)

6. Abundance Derivation

6.1 Curve of Growth

The relationship between equivalent width and elemental abundance has three regimes (see Figure 3, bottom left panel):

- Linear regime (EW < ~50 mÅ): EW ∝ N. Direct proportionality — most reliable.
- Flat part (EW ~50–150 mÅ): Line core is optically thick. EW grows slowly. Sensitive to ξ.
- Damping regime (EW > ~150 mÅ): Pressure-broadened wings dominate. EW ∝ √N.

We preferentially use lines in the linear regime and use Voigt profiles for strong lines in the damping regime.

6.2 Solar Normalization

All results are expressed on the [X/H] scale, normalized to the Sun using Asplund et al. (2021):

```
[X/H] = A(X)_star − A(X)_sun    (Asplund et al. 2021)
```

Important: Solar Reference Difference vs 2010

The 2010 predecessor study (Schmitt 2010) used Lodders (2003). The current standard is Asplund et al. (2021). The solar iron abundance changed: A(Fe)⊙ = 7.50 (Lodders 2003) → 7.46 (Asplund 2021). All comparisons to 2010 results must account for this 0.04 dex offset.

7. Uncertainty Analysis

The uncertainty budget follows the Type A / Type B framework of the JCGM Guide to the Expression of Uncertainty in Measurement (GUM) — the same framework used in precision metrology (e.g., laser power measurements).

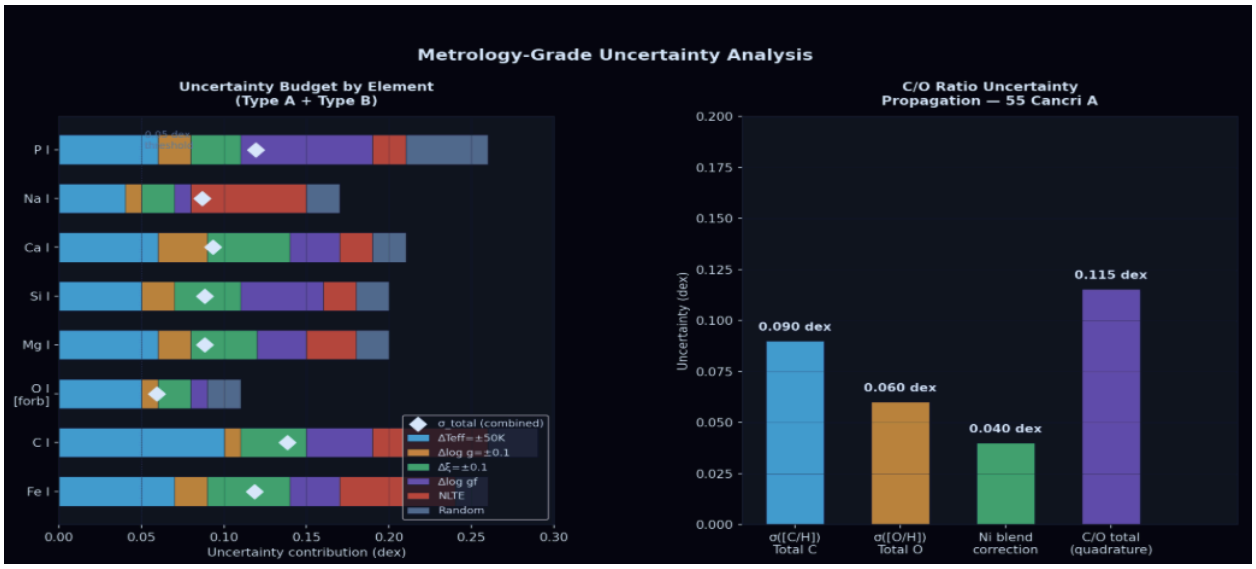

Figure 4. Metrology-grade uncertainty analysis. Left: Tornado chart showing contribution of each uncertainty source (Teff, log g, ξ, log gf quality, NLTE corrections, photon noise) to the total σ for each element. Diamonds show combined σ_total. Right: C/O ratio uncertainty propagation for 55 Cancri A, showing how carbon, oxygen, and the Ni blend correction combine in quadrature. The literature disagreement on C/O (Teske 2013 vs others) is indicated for context.

7.1 Type A Uncertainties (Random — reducible with more data)

Source	Estimation method	Typical magnitude
Photon noise on EW	Covariance matrix of Voigt/Gaussian fit	σ_EW ≈ 1.5 × FWHM / S/N
Line-to-line scatter	σ / √N for N lines of same element	~0.02 – 0.05 dex
Continuum placement	Repeat normalization with varied anchor points	~0.01 – 0.03 dex

7.2 Type B Uncertainties (Systematic — not reducible by more data)

Source	Estimation method	Typical magnitude
ΔTeff = ±50 K	Re-run abundances at Teff ± 50 K	±0.05 – 0.10 dex
Δlog g = ±0.1	Re-run at log g ± 0.1	±0.02 – 0.05 dex
Δξ = ±0.1 km/s	Re-run at ξ ± 0.1	±0.03 – 0.06 dex
log gf uncertainty	NIST grade (A: 3%, B: 10%)	~0.02 – 0.05 dex
NLTE corrections	Uncertainty on correction grids	~0.02 – 0.08 dex

Source	Estimation method	Typical magnitude
1D vs 3D models	Compare ATLAS9 vs MARCS results	~0.03 – 0.10 dex (C, O)

Combined uncertainty (in quadrature):

```
σ_total = √(σ_A² + σ_Teff² + σ_logg² + σ_ξ² + σ_loggf² + σ_NLTE²)
```

8. Pipeline Data Model

Figure 5. Pipeline data flow diagram. Three input sources (ESO Archive HARPS spectra, VALD3 line list, ATLAS9 model grid) feed into a 6-stage Python pipeline producing science outputs including C/O ratio, Mg/Si, Fe/Si, [α/Fe], and CHNOPS habitability index. All constants and parameters are loaded from config/constants.py — no hardcoded values in analysis scripts.

8.1 Key Output File Schemas

EW measurements table ({star_id}_ew_measurements.csv):

```
element, ion, wavelength_air_A, excitation_potential_eV, log_gf,nist_grade, EW_mA,
EW_err_mA, center_fitted_A, profile_used,chi2_reduced, blend_flag, quality_flag,
notes
```

Final abundances table ({star_id}_abundances_final.csv):

```
element, ion, n_lines, A_X_mean, A_X_scatter, A_X_solar_A2021,XH, XH_err_typeA,
XH_err_typeB, XH_err_total,XFe, XFe_err_total, nlte_correction, nlte_source
```

Science outputs table ({star_id}_science_outputs.csv):

```
star_id, teff_K, logg, feh, age_gyr,CO_ratio, CO_err, MgSi_ratio, MgSi_err,
FeSi_ratio, FeSi_err,alpha_Fe_mean, alpha_Fe_err, CHNOPS_index, CHNOPS_err
```

9. Solar System Validation

Before reporting any science results, the pipeline is validated against the Sun — the only star for which we know the true answer.

9.1 Solar Calibration Asteroid

HARPS observations of asteroids Vesta and Ceres provide reflected sunlight — a solar spectrum processed through exactly the same instrument and pipeline as the science targets. This is the most rigorous possible validation.

Validation criteria:

- A(Fe) = 7.46 ± 0.05 (Asplund et al. 2021)
- A(Na) = 6.24 ± 0.04 (Asplund et al. 2021)
- C/O = 0.55 ± 0.05 (solar ratio)

If any criterion fails, the pipeline has a systematic error that must be identified and corrected before any 55 Cancri results are reported.

12. References

Amarsi A.M. et al. (2019), A&A 630, A104 — Carbon NLTE corrections

Amarsi A.M. et al. (2021), A&A 656, A113 — Oxygen NLTE corrections

Amarsi A.M. et al. (2022), A&A 668, A68 — Iron NLTE corrections

Asplund M. et al. (2021), A&A 653, A141 — Solar abundances (current standard)

Barklem P.S. & Aspelund-Johansson J. (2005), A&A 435, 373 — Damping constants

Blackwell D.E. et al. (1979), MNRAS 186, 657 — Microturbulence method

Blanco-Cuaresma S. et al. (2014), A&A 569, A111 — iSpec

Castelli F. & Kurucz R.L. (2003), Proc. IAU Symp. 210 — ATLAS9 models

Fischer D.A. & Valenti J. (2005), ApJ 622, 1102 — 55 Cnc spectral parameters

Fischer D.A. et al. (2008), ApJ 675, 790 — 55 Cnc radial velocities

Gustafsson B. et al. (2008), A&A 486, 951 — MARCS models

Kramida A. et al. (2023), NIST Atomic Spectra Database — physics.nist.gov/asd

Lind K. et al. (2011), A&A 528, A103 — Sodium NLTE corrections

Lodders K. (2003), ApJ 591, 1220 — Solar abundances (used in Schmitt 2010)

Ryabchikova T. et al. (2015), PhyS 90, 054005 — VALD3

Schmitt R. (2010), Senior Thesis, University of Montana — predecessor study

Sneden C. (1973), ApJ 184, 839 — MOOG

Teske J.K. et al. (2013), ApJ 778, 132 — 55 Cnc C/O ratio

Valenti J.A. & Fischer D.A. (2005), ApJS 159, 141 — Spectral synthesis method

van Leeuwen F. (2007), A&A 474, 653 — Hipparcos parallax catalog

von Braun K. et al. (2011), ApJ 740, 49 — 55 Cnc CHARA interferometric parameters