The Astrophysical Journal, 1003:27 (16pp), 2026 May 20 https://doi.org/10.3847/1538-4357/ae6108

Observational Constraints on the Origin of the Elements. X. Combining Non–Local Thermodynamic Equilibrium and Machine Learning for Chemical Diagnostics of 4 Million Stars in the 4MIDABLE-HR Survey

Nicholas Storm[1,2], Maria Bergemann[1], Tomasz Różański[3], Victor F. Ksoll[4], Thomas Bensby[5], Gregor Traven[6], Georges Kordopatis[7], Ross P. Church[5], Mingjie Jian[8], Weijia Sun[9], Guillaume Guiglion[10,1], and Gražina Tautvaišienė[11]
[1] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany; storm@mpia.de
[2] Heidelberg University, Grabengasse 1, 69117 Heidelberg, Germany
[3] Research School of Astronomy & Astrophysics, The Australian National University, Cotter Rd., Weston, ACT 2611, Australia
[4] Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[5] Lund Observatory, Division of Astrophysics, Department of Physics, Lund University, Box 118, SE-221 00 Lund, Sweden
[6] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia
[7] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Nice, France
[8] Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
[9] Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
[10] Zentrum für Astronomie der Universität Heidelberg, Landessternwarte, Königstuhl 12, 69117 Heidelberg, Germany
[11] Vilnius University, Faculty of Physics, Institute of Theoretical Physics and Astronomy, Sauletekio av. 3, 10257 Vilnius, Lithuania
Received 2025 December 5; revised 2026 April 14; accepted 2026 April 16; published 2026 May 12

Abstract

We present the 4MOST-HR resolution non–local thermal equilibrium (NLTE) Payne artificial neural network (ANN), trained on 404,793 new FGK spectra with 16 elements computed in NLTE. This network will be part of the Stellar Abundances and atmospheric Parameters Pipeline (SAPP), which will analyze 4 million stars during the 5 yr long 4MOST consortium 4: 4MOST MIlky way Disc And BuLgE High-Resolution (4MIDABLE-HR) survey. A fitting algorithm using this ANN is also presented that is able to fully automatically and self-consistently derive both stellar parameters and elemental abundances. The ANN is validated by fitting 121 observed spectra of low-mass FGKM-type stars, including main-sequence dwarf, subgiant, and giant stars down to [Fe/H] ≈ -3.3 degraded to a 4MOST-HR resolution of $R \approx 20{,}000$ and comparing the derived abundances with the output of the classical radiative transfer code TSFitPy. We are able to recover all 18 elemental abundances with a bias of <0.13 and spread of <0.16 dex, although the typical values are <0.09 dex for most elements. These abundances are compared to the OMEGA+ Galactic chemical evolution model, showcasing for the first time the expected performance and results obtained from high-resolution spectra of the quality expected to be obtained with 4MOST. The expected Galactic trends are recovered, and we highlight the potential of using many chemical elements to constrain the formation history of the Galaxy.

Unified Astronomy Thesaurus concepts: Neural networks (1933); Stellar abundances (1577); Galaxy chemical evolution (580)

Materials only available in the online version of record: machine-readable tables

1. Introduction

Recent advances in spectroscopic facilities have enabled the collection of millions of individual stellar spectra in the Milky Way. One of them is the upcoming 4MOST facility with first light in late 2025. The 4MOST consortium survey 4: MIlky way Disk And BuLgE High-Resolution (4MIDABLE-HR) survey will provide high-resolution and high signal-to-noise ratio (SNR) spectra for over 4 million stars in the Galactic disk and the bulge during its 5 yr program (T. Bensby et al. 2019). It will precisely measure chemical abundances not only to constrain their origin and production sites but also to better understand the formation and history of our Galaxy (G. Gilmore et al. 1995; J. Bland-Hawthorn & O. Gerhard 2016; B. Barbuy et al. 2018; J. J. Cowan et al. 2021; A. Arcones & F.-K. Thielemann 2023). However, analysis of this enormous amount of data cannot be efficiently done on a timely scale using classical spectral synthesis codes such as MOOG (C. A. Sneden 1973), SIU (J. K. Reetz 1991), PySME (A. Wehrhahn et al. 2023), KORG (A. J. Wheeler et al. 2023), or TSFitPy (J. M. Gerber et al. 2023; N. Storm & M. Bergemann 2023). For this reason, a lot of spectroscopic surveys turned to different types of machine learning algorithms such as the data-driven approach Cannon (M. Ness et al. 2015), convolutional neural networks (CNNs; see, e.g., G. Guiglion et al. 2024), and the Payne (Y. S. Ting et al. 2019), a forward generative neural network for synthetic spectra emulation.

In this paper, we present a non–local thermodynamic equilibrium (NLTE) Payne artificial neural network (ANN) trained on 404,793 newly computed spectra in NLTE. Our main motivation is to provide a comprehensive and fast spectroscopic analysis code with state-of-the-art models computed—to the extent currently possible—using NLTE physics, which is required for accurate stellar parameter and chemical abundance analyses of stars (K. Lind & A. M. Amarsi 2024; M. Bergemann & R. Hoppe 2025). Our approach is based on a single network that emulates spectra with 16 elements simulated in NLTE across the full FGK

parameter space. This emulator-based approach provides fast spectral evaluations, introduces no explicit parameter priors, and integrates with standard fitting strategies that rely on features less sensitive to spectral-model systematics (e.g., line-list imperfections or blends). Moreover, emulation provides direct access to the synthetic spectrum at the best-fitting parameters, facilitating manual inspection of residuals and overall fit quality. While the NLTE Payne is not new (e.g., M. Kovalev et al. 2019; S. Buder et al. 2025), we are not aware of any prior single-network high-resolution NLTE Payne ANN trained over such an extensive parameter space, with NLTE effects taken into account for 15+ elements. However, we note similar works by S. Buder et al. (2025), which contained 14 elements in NLTE but split the parameter space over many dozens of small neural networks, and works by M. Xiang et al. (2019) and M. Zhang et al. (2025), who utilized a slightly different architecture, "DD-Payne," and applied it to lower-resolution LAMOST spectra. Our network is planned to be part of the Stellar Abundances and atmospheric Parameters Pipeline (SAPP) in 4MOST (for the 4MIDABLE-HR survey) and the PLATO missions (M. R. Gent et al. 2022; K. Lee et al. 2026, in preparation), but it can be used alone as well. We also release this 4MOST-HR network for public use, together with 4MOST-LR and a high-resolution version ($R = 80{,}000$),[12] as well as the code used for Payne training.[13] We validate our new NLTE ANN abundances by comparing them with the results obtained using the classical radiative transfer code TSFitPy (J. M. Gerber et al. 2023; N. Storm & M. Bergemann 2023). We also compare our self-consistent NLTE [X/Fe] trends with the predictions of our well-established Galactic chemical evolution (GCE) model OMEGA+ (B. Côté et al. 2017, 2018) to highlight the parameter space of chemical enrichment studies, as planned with the 4MIDABLE-HR survey, but also to expose the current limitations and potential of such comparative analyses.

The paper is organized as follows. We present our NLTE spectral synthetic models, trained NLTE Payne ANN, observed stellar sample, and GCE model in Section 2. We validate the derived stellar parameters with the literature and the abundances with results from a classical fitting method with TSFitPy and compare to the GCE models to constrain star formation history (SFH) in Section 3. Finally, we briefly summarize the conclusions in Section 4.

2. Data and Method

2.1. NLTE Payne Model

We adopt the Payne architecture (Y. S. Ting et al. 2019), following its NLTE adaptation by M. Kovalev et al. (2019). Recent scaling results (T. Różański & Y.-S. Ting 2025) indicate that emulation accuracy improves predictably with training-set size, computing time spent for training, and neural network size; thus, the achievable precision is set by resource and latency constraints. In this work, we require a full per-spectrum fit (stellar parameters and 17 elemental abundances) to be completed within around 10–20 s on a personal computer not equipped with a GPU. This requirement is set such that we are able to fit spectra of several hundred (up to several thousand) stars at high resolution for each night of 4MOST observations. Under this constraint, a compact Payne multilayer perceptron (MLP) offers a better accuracy-latency trade-off than heavier alternatives (e.g., TransformerPayne; T. Różański et al. 2025), especially considering the fact that we can afford the computation of a spectral grid with hundreds of thousands of spectra.

Our NLTE Payne is a four-layer MLP. The input layer ingests 21 parameters: four atmospheric stellar parameters (effective temperature $T_{\rm eff}$, surface gravity $\log g$, metallicity [Fe/H], and microturbulent velocity parameter $\xi_{\rm t}$) and the abundances that are core for 4MIDABLE-HR survey science (T. Bensby et al. 2019; M. Bergemann et al. 2026, in preparation) of Li, C, O, Na, Mg, Al, Si, Ca, Ti, Cr, Mn, Co, Ni, Sr, Y, Ba, and Eu. Three fully connected hidden layers with 1024 units each and sigmoid linear unit (SiLU) activations transform the inputs, and a final output layer, equipped with a sigmoid activation function, predicts continuum-normalized flux at 33,375 wavelength pixels. This is the largest model that still satisfies our latency constraint. We have opted for SiLU in the middle layers and sigmoid in the output layer activation functions because, according to our tests, these functions results in lower final losses of the network. Further training details, the test results testing different activation functions, and the hyperparameter optimization strategy are provided in Appendix A.

2.2. NLTE Synthetic Spectra

We trained the NLTE Payne described above on a large, newly computed NLTE synthetic spectral grid with 404,793 spectra, with generation requiring roughly 125,000 CPU hr (see Figure 1 for coverage), which also makes it one of the largest grids ever computed. The grid covers the 4MOST-HR blue (3926–4355 Å), green (5160–5730 Å), and red (6100–6790 Å) windows (R. S. de Jong et al. 2019). We created these spectra using the models and methods developed in our previous work (M. Bergemann et al. 2019, 2021; E. Magg et al. 2022; J. M. Gerber et al. 2023). We used the standard 1D MARCS model atmospheres (B. Gustafsson et al. 2008) and NLTE departure grids computed using our well-tested and validated NLTE atomic models of H (L. Mashonkina et al. 2008), O (M. Bergemann et al. 2021), Na (R. Ezzeddine et al. 2018), Mg (M. Bergemann et al. 2017), Al (R. Ezzeddine et al. 2018), Si (M. Bergemann et al. 2013; E. Magg et al. 2022), Ca (L. Mashonkina et al. 2017; E. Semenova et al. 2020), Ti (M. Bergemann 2011), Mn (M. Bergemann et al. 2019), Fe (M. Bergemann et al. 2012b; E. Semenova et al. 2020), Co (M. Bergemann et al. 2010; S. A. Yakovleva et al. 2020), Ni (M. Bergemann et al. 2021; Y. V. Voronov et al. 2022), Sr (M. Bergemann et al. 2012a; J. M. Gerber et al. 2023), Y (N. Storm & M. Bergemann 2023; N. Storm et al. 2024), Ba (A. J. Gallagher et al. 2020), and Eu (N. Storm et al. 2024). The MARCS atmosphere grid relies on the chemical composition adopted from N. Grevesse & A. J. Sauval (1998) and N. Grevesse et al. (2007). However, in our NLTE synthetic spectrum calculations, we adopt our recent self-consistence solar composition from E. Magg et al. (2022), which was obtained using Turbospectrum v.20 in NLTE. In Appendix C, we demonstrate how the new NLTE Payne performs on the 4MOSTified solar spectra, and we then use the new solar abundances in the analysis of GCE results in Section 3.3.

Our grid coverage is presented in Figure 1 and covers the following ranges: $3500 \leqslant T_{\rm eff}/{\rm K} \leqslant 8000$, $0.5 \leqslant \log g \leqslant 5.0$, $-5 \leqslant {\rm [Fe/H]} \leqslant 0.5$, and $0.5 \leqslant \xi_{\rm t}/{\rm km\ s^{-1}} \leqslant 5$. Individual

[12] https://nlte.mpia.de/gui-payne_fit.php
[13] https://github.com/stormnick/one_payne

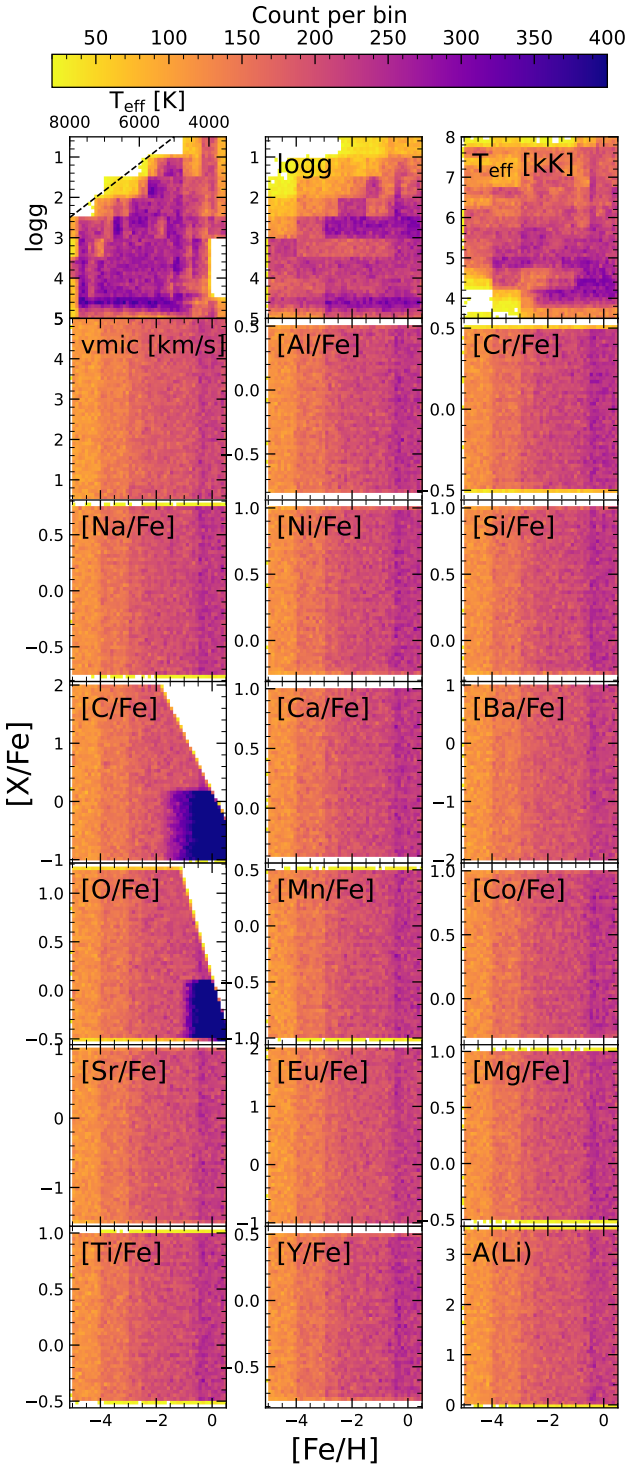

Figure 1. Distribution in a 2D histogram of synthetic spectra used to train the Payne. The top left panel shows the Kiel diagram, while the rest are distributions as a function of [Fe/H]. All abundances are chosen to be uniformly random in metallicity space, except for A(O) < 8.87 and A(C) < 8.7. There are fewer low-metallicity giant model atmospheres, resulting in slightly fewer spectra at low metallicities. There are also no public MARCS models above the black line in the $T_{\rm eff}$–$\log g$ space, resulting in a lack of computed spectra in that regime.

elements have distinct ranges that are slightly larger than the typically observed values. For red giants, there are no MARCS models at low metallicity, resulting in fewer spectra in that parameter space. Most importantly, the individual abundances were chosen to be uniformly random—there is no training bias (i.e., expected abundance pattern) that goes into the ANN. Only for C and O did we choose values such that A(O) < 8.87 and A(C) < 8.7 to remove any spectra with a very high abundance of C and O at high metallicity, causing unrealistic and too-strong molecular bands in the training set. This also means that our network can effectively measure abundance in nonstandard chemical composition stars. We adopted the Gaia-ESO line lists (U. Heiter et al. 2021) with new atomic data for C, N, O, Si, and Mg, as described in E. Magg et al. (2022), and used VALD for its gaps (T. Ryabchikova et al. 2015). We also adopt solar abundances from E. Magg et al. (2022). The NLTE Turbospectrum and TSFitPy wrappers (J. M. Gerber et al. 2023; N. Storm & M. Bergemann 2023) were employed to compute the spectra at infinite resolving power, which were further degraded to $R \approx 20{,}000$ ("4MOSTified") using the 4FS ETC package.[14] In order to simplify the Payne architecture, we have opted to not apply any macroscopic broadening to the synthetic templates. Instead, we apply the convolution with a broadening profile after the neural network inference and during the χ^2 minimization parameter optimization. Even at the resolving power of the 4MOST high-resolution spectrograph, it is not possible to disentangle different components, such as $V_{\rm mac}$ and $V_{\rm sini}$, from the shape of the spectral lines, which is why we only add a convolution with the latter profile.[15] This implies that the $v_{\rm brd}$ estimates represent the combined broadening due to projected (for unknown inclination) rotational velocity and macroturbulence, but since the latter is an ad hoc parameter introduced in 1D hydrostatic equilibrium models to compensate for the absence of realistic convection and turbulence (M. Asplund 2005; Å. Nordlund et al. 2009; K. Lind & A. M. Amarsi 2024), we do not optimize this parameter further. This approach allows us to keep the Payne network loss low while still allowing flexible fitting of rapidly rotating stars.

2.3. Spectroscopic Analysis

Although the fitting methodology is not the main focus of this work, as it is similar to the classical synthetic codes, we briefly describe our method. We used the `minimize` function from Python package `LMFIT` (M. Newville et al. 2025) to fit the spectra, which takes on the order of 10–20 s to fit one spectrum (both stellar parameters and all 17 elemental abundances) on a single Mac M2 CPU core.

The four atmospheric parameters ($T_{\rm eff}$, $\log g$, [Fe/H], $\xi_{\rm t}$) and broadening ($v_{\rm brd}$) were determined by simultaneously fitting Fe I, Fe II, Mg I, and Ca I lines (Table D1). We did not impose priors on abundance ratios during the fit, nor did we adopt any scaling functions, such as, e.g., the commonly used [α/Fe] relation with metallicity. Fitting Fe lines enforces ionization balance, while Mg and Ca lines add additional constraints. We decided against usage of hydrogen lines, since they are badly modeled in giants, which gives poor results for those stars (see, e.g., S. Wedemeyer et al. 2017). The final best-fit values are then used to determine abundances by sequentially fitting the abundances of the remaining individual elements. Abundances of all 18 chemical elements (including Fe) were derived by minimizing the χ^2 in the spectral regions contained by the line

[14] https://escience.aip.de/readthedocs/OpSys/etc/master/index.html

[15] See here: https://github.com/stormnick/one_payne/blob/main/src/convolve.py.

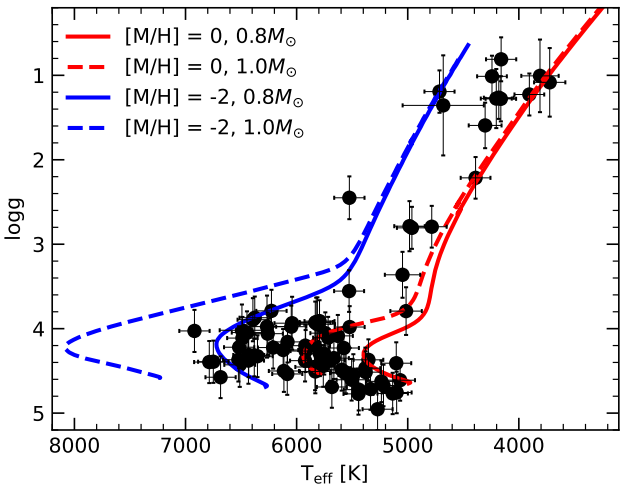

Figure 2. Kiel diagram of the fitted stellar sample with PARSEC evolutionary tracks (A. Bressan et al. 2012) in color, with red tracks having [M/H] = 0, blue [M/H] = -2, solid mass of 0.8 $M_\odot$ and dashed mass of 1 $M_\odot$.

masks centered on the diagnostic lines. For Eu, we subtract 0.31 dex from all derived abundances, because in the standard line list available to us, the log gf was offset by a constant factor of +0.31 (log(2)) compared to the accurate experimental f-value for this transition from J. E. Lawler et al. (2001). The list and the atomic parameters of the fitted lines are provided in Table D1. Fe lines were chosen from the selection done based on G. Kordopatis et al. (2023).[16] However, we note that the line masks and the fitting algorithm are not fixed, and both can be easily adjusted without changing the NLTE ANN itself.

2.4. Stellar Sample

Our calibration stars were selected from a sample of benchmark stars (U. Heiter et al. 2015), and we also included spectra of nearby bright stars from our previous studies in K. Fuhrmann et al. (1993), K. Fuhrmann (1998), T. Gehren et al. (2004, 2006), and M. Bergemann & T. Gehren (2008), totaling 121 individual spectra for 84 stars. The coverage of the stellar parameters is plotted in Figure 2 and covers the FGKM space in the ranges $3700 \lesssim T_{\rm eff}/{\rm K} \lesssim 6900$, $0.8 \lesssim \log g \lesssim 5.0$, and $-3.35 \lesssim$ [Fe/H] $\lesssim 0.34$. Our calibration sample consists of 48 main-sequence stars, 19 subgiants, and 17 red giants. The observed spectra for the benchmark stars were taken from S. Blanco-Cuaresma et al. (2014), and they include high-resolution HARPS, NARVAL, and UVES spectra taken at an SNR of typically better than 200 Å^{-1}. The wavelength coverage varies, and for the purpose of this work, we trimmed it to the 4MOST-HR windows. The rest of the sample was taken using FOCES, and the spectra have $R \approx 40{,}000$, a typical SNR of at least 200 Å^{-1}, and a wavelength range of 4500–6850 Å (with a few spectra extending from 3930 to 7450 Å). Some of our spectra do not cover the blue 4MOST-HR window.

We degraded the spectra to a resolution of $R \approx 20{,}000$ using a simple fast Fourier transform convolution. It is important to note that the actual 4MOST spectra are expected to have a resolution that varies with wavelength (from 18,000 in the shorter-wavelength regime to 21,000 in the longer-wavelength regime for each spectral window; see R. S. de Jong et al. 2019; C. J. Walcher et al. 2019).[17] Thus, we compared all derived Payne abundances with the ones fitted using TSFitPy. This is the only reliable way to do a self-consistent analysis and showcase the capabilities of the NLTE ANN itself without differences coming from using distinct spectra, model atmospheres, atomic data, spectra resolution, etc., if compared to the literature values.

2.5. GCE Model

To compute the chemical evolution of the elements discussed in this paper, we used the OMEGA+ GCE code (B. Côté et al. 2017, 2018). This is a standard analytical GCE model, and it tracks the galactic gas mass, increasing with the inflow from the circumgalactic medium (CGM) and the stellar mass-loss rate and decreasing due to star formation and outflows. The code also tracks the elemental abundances in the interstellar medium (ISM) as a function of time, providing synthetic data in the form of [X/Fe] that can be directly compared to observational data. The values of the input parameters in OMEGA+ are provided in Table 1. The choices of these values are discussed and justified in J. Lian et al. (2023), but we summarize them briefly here. We used an exponential inflow rate, which adds gas regardless of the available amount in CGM,

$$\dot{M}_{\rm inflow}(t) = N_{\rm norm} e^{-t/\tau}, \tag{1}$$

where $\dot{M}_{\rm inflow}$ is the inflow gas, t is the age of the Galaxy in Gyr, $N_{\rm norm}$ is the normalization constant (chosen as 10 $M_\odot\,{\rm yr}^{-1}$ here), and τ is the infall timescale (chosen as 7 Gyr here). It is important to note that OMEGA+ tracks the amount of gas in the CGM ($M_{\rm CGM}$) with the initial amount at $t = 0$ defined as (see also Equation (4) in B. Côté et al. 2018)

$$M_{\rm CGM}(t=0) = \frac{\Omega_{b,0}}{\Omega_0} M_{\rm vir}, \tag{2}$$

with adopted cosmological parameters $\Omega_{b,0} = 0.05$ and $\Omega_0 = 0.32$, where $M_{\rm vir}$ includes both dark matter (DM) and baryonic mass (here initial $M_{\rm vir} \equiv M_{\rm DM}$, where $M_{\rm DM}$ is constant at all time steps). If at any time step, the gas in the CGM $M_{\rm CGM}$ runs out due to galactic inflows, then Equation (1) is not used during that time step for the inflow calculations; instead, a small constant inflow rate is calculated,

$$\dot{M}_{\rm inflow}(t) = \begin{cases} \dot{M}_{\rm calc}, & \text{if} \dot{M}_{\rm calc}\ \Delta t \leqslant M_{\rm CGM}, \\ M_{\rm CGM}/\Delta t, & \text{if} \dot{M}_{\rm calc}\ \Delta t > M_{\rm CGM}, \end{cases} \tag{3}$$

where Δt is the size of the time step in years and $\dot{M}_{\rm calc}$ is the calculated inflow rate from Equation (1). Therefore, if the initial CGM gas mass is too low or the infall rate is too large, star formation will be quickly suppressed, as little new gas is added to the galaxy. However, as long as the CGM has enough gas, the initial DM mass does not affect the inflow rate (which is the case for our galaxy simulation here). In OMEGA+, the star formation efficiency (SFE) $\epsilon_\star$ is a dimensionless parameter that goes into the equation

$$\dot{M}_\star = \frac{\epsilon_\star}{\tau_\star} M_{\rm gas}, \tag{4}$$

[16] https://line-detector.oca.eu

[17] See also www.4most.eu.

Table 1
Chosen Parameters for the OMEGA+ GCE Model

Parameter	Chosen/Used Values	Comment/Yield Source
Inflow rate	$\dot{M}_{\rm inflow}(t) = N_{\rm norm}\, e^{-t/\tau}$	$N_{\rm norm} = 10\ M_\odot\ {\rm yr}^{-1}$, $\tau = 7$ Gyr
$M_{\rm DM}$	$10^{12}\ M_\odot$	Constant at all time steps
SFE $\epsilon_\star$	0.03	Our default GCE model
Mass loading η	1	Used in $\dot{M}_{\rm out} = \eta \dot{M}_\star$
NSM	DTD(τ) $\propto \tau^{-1}$, 10 Myr $< \tau <$ 10^6 Gyr	M. Arnould et al. (2007)
CCSN	13–30 $M_\odot$	M. Limongi & A. Chieffi (2018)
MRSN	0.01% of 13–25 $M_\odot$ CCSN	N. Nishimura et al. (2015)
SN Ia	Four channels	P. Eitner et al. (2023)
AGB	1–6 $M_\odot$	S. Cristallo et al. (2015)

Note. All other parameters were kept as default.

where $\dot{M}_\star$ is the amount of gas going into the star formation, $M_{\rm gas}$ is the amount of available gas, and $\tau_\star$ is the star formation timescale. In our case, we keep the default configuration of $\tau_\star$; i.e., it is the dynamical timescale of the whole virialized system,

$$\tau_\star = 0.1 f_{\rm dyn}\, H_0^{-1} (1+z)^{-3/2}, \quad (5)$$

where H_0 is the Hubble constant, z is the redshift, and $f_{\rm dyn} = 0.1$ by default (see Equation (12) in B. Côté et al. 2017 and references therein). $\tau_\star$ starts at ≈6.7 Myr and goes up to ≈145.7 Myr at the end of the simulation. In other words, varying the dimensionless SFE parameter $\epsilon_\star$ proportionally affects the timescale at which the star formation is happening.

Our GCE model relies on yields for core-collapse supernovae (CCSNe) from M. Limongi & A. Chieffi (recommended set "*R*" 2018), which uses yields from different rotations based on the velocity distribution weights derived by N. Prantzos et al. (2018). For asymptotic giant branch (AGB) stars, we adopted the S. Cristallo et al. (2015) yields. We use the same approach as P. Eitner et al. (2023) for Type Ia supernova (SN Ia) yields with four channels, each having its own delay time distribution (DTD): single degenerate Chandrasekhar mass SNe Ia with H transfer from the companion via stable Roche lobe overflow, fainter SNe Iax, sub-$M_{\rm ch}$ with double detonation of a C-O white dwarf (WD), and sub-$M_{\rm ch}$ SN Ia due to a merger of two WDs. For the *r*-process, we utilize two sites. The first are neutron star mergers (NSM) with a DTD slope τ^{-1}, with an initial time delay of 10 Myr and a maximum one until 10^6 Gyr, according to the maximum merging time of neutron star–neutron star (NS-NS) systems (P. Beniamini & T. Piran 2019). The second are magnetorotational supernovae (MRSNe), where we used the same approach as C. Kobayashi et al. (2020) by replacing 0.01% of the CCSN yields in the range 13–25 $M_\odot$ with yields of a 25 $M_\odot$ MRSN from N. Nishimura et al. (2015).

3. Results

In Figure 3, we show two examples of the NLTE ANN fit of a 4MOSTified spectra of the benchmark metal-poor stars HD 140283 and HD 84937. Overall, the spectra are well reproduced. The rms error (RMSE) of the flux is ≈0.010 in normalized flux units over all three windows for both spectra, which is fully sufficient for our purposes, as the error is similar to the noise level of the data and other observational artifacts, such as data reduction and calibration, that introduce an even larger error.

3.1. Stellar Parameter Comparison

We validated the stellar parameters $T_{\rm eff}$ and $\log g$ by comparing our best-fit values to the values from the literature (U. Heiter et al. 2015). This comparison has revealed a small systematic bias of ≈−0.3 dex in the spectroscopic $\log g$ values. This is not surprising: a similar bias was also noted in the analyses based on the previous NLTE Payne version (see Section 4.1 in M. R. Gent et al. 2022). This bias also affects the derived $T_{\rm eff}$ and [Fe/H] values, because of the degeneracy in the fitting of stellar parameters. The degeneracy can be broken only by using other nonspectroscopic information, such as photometric and parallax data (e.g., R. Schönrich & M. Bergemann 2014; S. Buder et al. 2025). The SAPP version in 4MOST will use this information (see also A. M. Serenelli et al. 2013; R. Schönrich & M. Bergemann 2014). Presently, in order to mitigate the bias in the current results, we apply a constant bias correction of +0.3 dex to $\log g$, and then we redetermine all other stellar parameters and abundances.

Figure 4 compares the literature stellar parameters with our NLTE ANN results for the benchmark stars. As mentioned above, the literature values were adopted from U. Heiter et al. (2015), and these latter quantities rely on asteroseismic constraints for surface gravity and interferometric constraints on stellar angular diameter for $T_{\rm eff}$. The agreement between our values and the independent parameters is excellent; for $T_{\rm eff}$ and $\log g$, the average difference is ∼9 K and 0.06 dex, respectively. The standard deviation between the two distributions is 131 K and 0.24 dex. We note that using alternative sources of stellar parameters from C. Soubiran et al. (2024) results in similar differences: average differences of $T_{\rm eff}$ of 59 K and $\log g$ of 0.05 dex and standard deviations of 105 K and 0.19 dex. This agreement is fully satisfactory, also given the fact that the literature values are not free of systematic errors.

3.2. Abundance Comparison

The comparison between the absolute fitted abundance from Payne and TSFitPy is plotted in Figure 5. Most spectra did not have blue window coverage; thus, some elements were impossible to fit, especially at low metallicity. For all elements, both bias and spread are respectively less than 0.13 and 0.16 dex and <0.09 dex for most elements. Lithium, α-, and Fe-peak elements perform the best, given that they have the strongest and typically less blended features. We mention some other elements below.

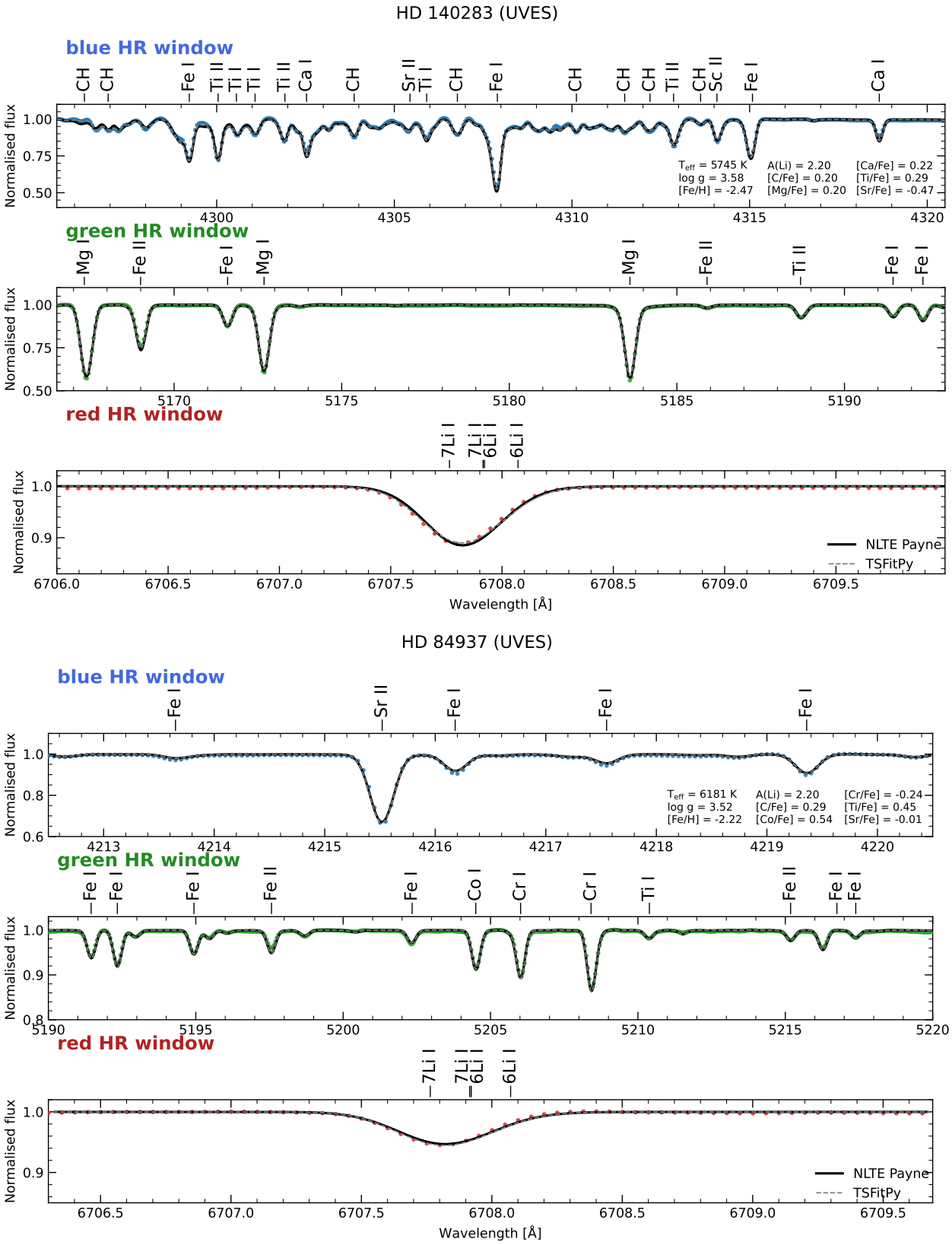

Figure 3. NLTE Payne fit (black lines) to the HD 140283 and HD 84937 UVES spectra (dots), degraded to $R \approx 20{,}000$ resolution, in all three 4MOST-HR windows. An emulated TSFitPy synthetic spectrum (gray dashed line) using the best-fit abundances is overplotted and nearly perfectly matches the Payne fit.

The star-to-star scatter of abundances of carbon (C) as derived using the NLTE Payne is much large compared to the values obtained with the TSFitPy code. We trace this back to the lack of the CH *G* band (≈4300 Å) in our observed 4MOSTified archival spectra. In the green, there is an atomic C I line at 5380 Å and C_2 molecular features at 5164 and

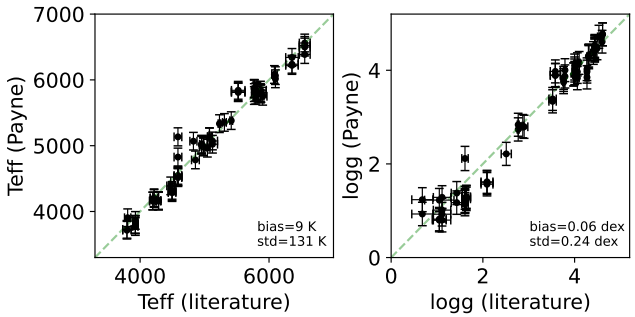

Figure 4. Comparison between the literature stellar parameters of $T_{\rm eff}$ and $\log g$ for benchmark stars from U. Heiter et al. (2015) to the Payne fitted ones.

5633 Å that are weak, and they are typically not visible at [Fe/H] $\lesssim -1$. Because the 4MOST-HR spectra will have a much wider spectral range, covering the CH *G* band, we do not expect such problems with the 4MIDABLE-HR data. The strongest atomic line of forbidden O at 6300 Å is severely blended with the telluric features, which in practice limits the use of this feature in the analysis of stellar O abundance (E. Pancino et al. 2017). Some chemical elements only have a few (<five) diagnostic lines that may still be blended. Strontium has prominent diagnostic lines in the blue spectral range at 4077 (Sr I), 4161 (Sr I), 4215 (Sr II), and 4305 (Sr II) Å. The Sr I line at 4077 Å is also blended with Fe I, which contributes to the uncertainty. Y II has only a few lines in the green window, and all of them are weak, resulting in a large spread of ≈0.16 dex. Also, the key *r*-process element Eu II is represented by a weak line in the red window at 6645 Å, typically used at higher [Fe/H], and there are two strong (yet blended) Eu II lines at 4129 and 4205 Å in the blue, which are only useful at lower metallicity. Therefore, our estimates of stellar Eu abundances typically have a larger uncertainty of up to 0.15 dex.

We estimate the expected systematic errors by taking the standard deviation between the Payne and TSFitPy abundances. For $T_{\rm eff}$ and $\log g$, on the other hand, we use the differences with respect to the U. Heiter et al. (2015) results for the benchmark stars. Figure 6 summarizes both results. This systematic abundance error can be considered a conservative estimate, as the addition of the blue window would improve the determination for some elements. The stellar parameter estimates are also expected to improve when derived as part of the full SAPP. Nevertheless, for all abundance values, we provide an error that is a flat sum of the systematic and statistical errors returned by the `LMFIT-minimize` function. In the subsequent discussion (Section 3.3), we primarily focus on the abundances on the square bracket notation [X/Fe]. In these calculations, we use our self-consistent solar NLTE ANN abundances obtained using the 4MOSTified solar spectra (Table C1). Chemical abundances obtained from different spectra of the same star are averaged.

In summary, we show that robust abundances are recovered by the NLTE Payne, and the values agree well with those obtained using the classical spectroscopic analysis methods.

3.3. Galactic Chemical Evolution Models

Studies of detailed chemical abundance patterns enable quantitatively probing the SFH of a stellar population (e.g., K. A. Venn et al. 2004; E. Tolstoy et al. 2009; M. Bergemann et al. 2017; F. Matteucci 2021). Because stars of different mass have different lifetimes, and their yields depend on mass and metallicity, the relative timing of CCSNe and SNe Ia leaves an imprint in the planes of α elements versus metallicity. But these differences are also readily visible in the chemical enrichment tracks of other chemical elements, enabling a more detailed set of constraints on the relative role of different events in the GCE. In this section, we focus solely on the SFE parameter as a proxy for the SFH, a key quantity in studies of galaxies (B. Barbuy et al. 2018; R. Maiolino & F. Mannucci 2019), but we refer to C. Conroy (2013) and R. S. Somerville & R. Davé (2015) for other types of constraints pertinent to Galaxy growth and formation history.

Figure 7 shows our new NLTE Payne [X/Fe] abundances in comparison with the OMEGA+ GCE model computed using different values of SFE $\epsilon_\star$: 0.01, 0.03 (our standard value; see J. Lian et al. 2023), and 1. All other GCE parameters were kept at their default values as in our previous work (see Table 1 for the chosen parameters and also J. Lian et al. 2023). Following V. Hegedüs et al. (2025), we rescale the GCE trend for Mg by 0.3 dex upward, in order to cure the misfit at low metallicity. For most chemical elements, our [X/Fe] distributions are in excellent agreement with the literature using NLTE (G. Zhao et al. 2016; T. Mishenina et al. 2017; N. Storm et al. 2025). Especially the α elements, such as Mg, Si, Ti, and Ca, follow the expected [Fe/H]−[α/Fe] trends (T. Bensby et al. 2014). The higher value of the SFE shortens the chemical enrichment timescale, increasing the contribution of massive stars to the chemical evolution, which in turn implies that SNe Ia and AGBs start contributing to the enrichment at higher [Fe/H] (with a higher [Fe/H] achieved at a given time step due to a larger contribution of massive stars). As a consequence, the [α/Fe] knee is shifted upward. Adopting the SFE of 1 leads to higher—compared to observations—values of [Mg/Fe] and [Si/Fe] at a low metallicity of [Fe/H] $\lesssim -2$. This difference potentially implies that lower SFE values would be preferred for the observed stellar population. Values of SFE lower than 0.03 do not yield any significant (here, "significant" implies "distinguishable" given the uncertainty of the chemical abundance measurements) differences in the [Fe/H]–[X/Fe] plane, except for Mn at [Fe/H] $\lesssim -3$.

The proton (p)-rich elements (T. Gehren et al. 2004, 2006) Na and Al are produced in massive stars (in the Ne–Na and Mg–Al cycles) and in AGB stars (A. I. Karakas & J. C. Lattanzio 2003; E. Tolstoy et al. 2009). In our GCE model, an increased value of SFE leads to increased ratios of [Na/Fe] and [Al/Fe] (by ∼0.2–0.3 dex) at [Fe/H] $\lesssim -1.5$. Also, the upturn of [Na/Fe] and [Al/Fe] as a function of [Fe/H] at [Fe/H] ≈ -1 is due to the contribution from AGBs, as also seen in AGB models (M. Forestini & C. Charbonnel 1997; N. Mowlavi 1999). The trends for [Na/Fe] and [Al/Fe] are similar to those presented by T. Bensby et al. (2014), where they applied NLTE correction to Na abundances.

The abundance ratios of iron-peak elements are higher compared to GCE model predictions, especially at low metallicity, which is consistent with our previous NLTE results for Mn, Co, and Ni in 3D NLTE (N. Storm et al. 2025). The trend of [Mn/Fe] remains slightly subsolar, whereas [Co/Fe] shows a mild increase toward lower metallicity, consistent with M. Bergemann et al. (2010). For Ni and Co, our observed sample is limited to [Fe/H] $\gtrsim -1.5$. For comparison, the LTE values of [Cr/Fe] and

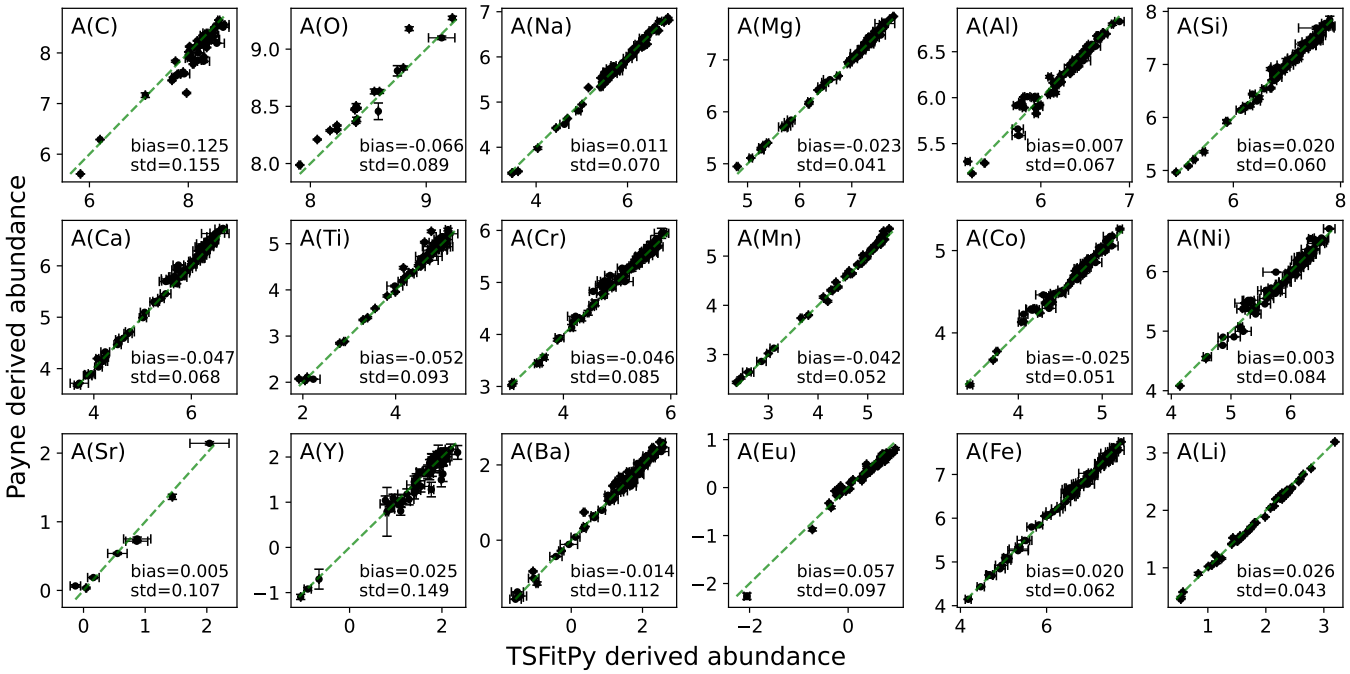

Figure 5. Comparison between the derived abundance from Payne and TSFitPy plotted in absolute A(X) units. The green line is a one-to-one comparison. Each subpanel shows the average difference (bias) and standard deviation (std) when comparing the abundances from the two sources.

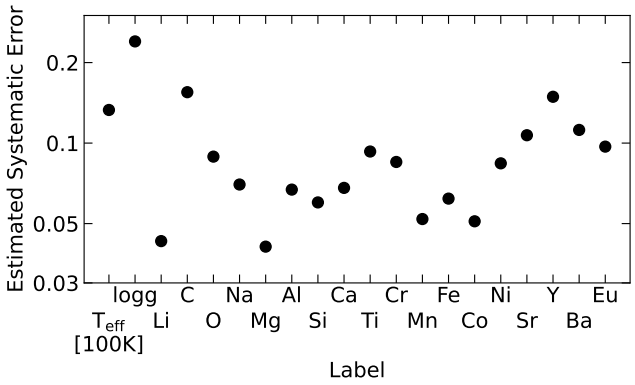

Figure 6. Estimated systematic error for stellar parameters and abundances. The error was estimated by taking the spread of the difference between the derived parameter from Payne and the literature (for $T_{\rm eff}$ and $\log g$ for benchmark stars from U. Heiter et al. (2015) or TSFitPy.

[Ni/Fe] presented by T. Bensby et al. (2014) follow a similar flat trend with [Fe/H]. Similarly, the NLTE [Mn/Fe] and [Co/Fe] values of C. Battistini & T. Bensby (2015) resemble our results. However, we note that the latter paper relies on our older[18] NLTE model atom of Mn and Co; therefore, the direct comparison is not meaningful.

Neutron-capture elements, such as Sr, Y, and Ba, have significant *r*-process and weak *s*-process contributions from massive stars at the lowest metallicities. At higher metallicities, on the other hand, the main *s*-process has a contribution only after the time delay of the formation of low- and intermediate-mass AGB stars. Thus, the metallicity at which the main *s*-process contributes gives an additional constraint on the SFH. Our GCE model for the *s*-process element tracers [Sr, Y, Ba/Fe] $\lesssim 0$ at [Fe/H] $\lesssim -1.5$ exhibits a significant AGB rise at [Fe/H] $\gtrsim -2$. An SFE of 1 results in abundance ratios 0.2–0.6 dex lower compared to a higher SFE at a given metallicity for [Fe/H] < -1. Thus, the upturn in *s*-process offers insight into the onset of AGB enrichment and star formation timescale. Trends of ratios of [Y/Fe] and [Ba/Fe] reported by T. Bensby et al. (2014) are flat with [Fe/H], but some stars have supersolar [Ba/Fe] at solar [Fe/H]. We notice a similar pattern for our stars. For [Eu/Fe], we note an upturn toward the lower [Fe/H], similar to our GCE trend and the expected Galactic trend seen in C. Battistini & T. Bensby (2016).

In summary, we show that the [X/Fe] obtained with the NLTE Payne are overall in agreement with the results reported in the literature. Also, the comparison of [Fe/H]–[X/Fe] with the predictions of the OMEGA+ GCE model suggests that the abundances inferred from spectra are overall consistent with the expectations for the chemical enrichment of the Galactic disk. These NLTE trends provide a preview of the trends expected from the 4MIDABLE-HR survey, which will be based on real 4MOST spectroscopic observations.

3.4. Lithium

We discuss Li in a separate section, as we have not previously tested our OMEGA+ GCE model for Li. Instead, we opt to use the B. D. Fields & K. A. Olive (1999a, 1999b) models, which include several different production channels: primordial ^{7}Li, production from Galactic cosmic-ray (GCR) reactions, and production from the SN ν-process.

Figure 8 demonstrates A(Li) as a function of [Fe/H], with our derived abundances shown as black points and lines corresponding to different GCE distributions. For [Fe/H] $\lesssim -1.6$, our stars lie on top of the curve with a scatter of $\approx$0.2, which means that our derived Li abundance is consistent with the GCE model. Similar to the results of S. G. Ryan et al. (2000), we

[18] These NLTE models made use of approximate classical rate coefficients describing the collisionally induced processes in H+Co and H+Mn transitions. In the later versions, these values were substituted by accurate quantum-mechanical values.

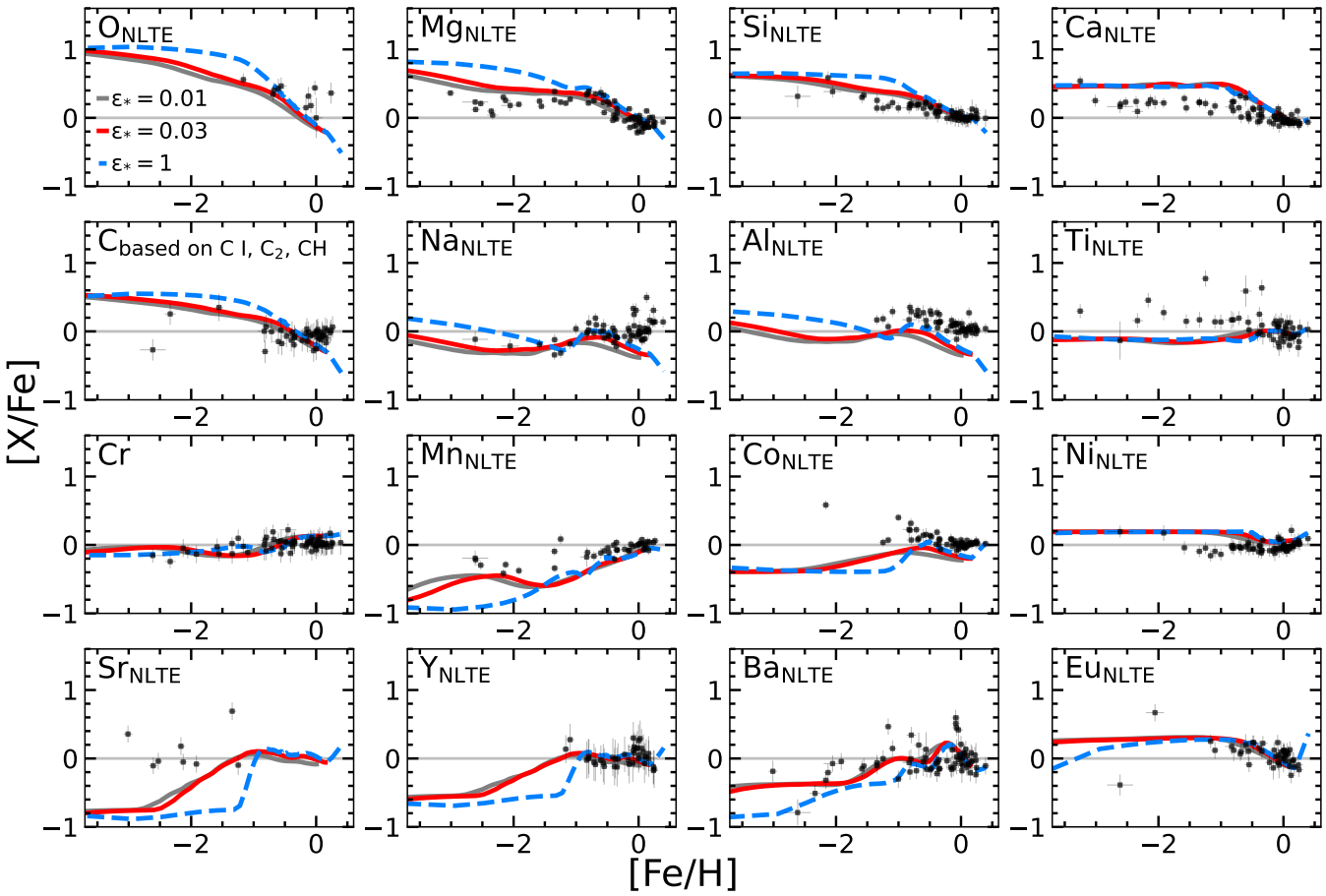

Figure 7. Chemical abundances derived using the new NLTE Payne ANN (black points) from the observed archival data of the Gaia-ESO benchmark and nearby bright stars. The archival spectra are described in Section 2.4. OMEGA+ GCE models are overplotted with three distinct SFEs $\epsilon_\star$: 0.01 (solid gray), 0.03 (solid red), and 1 (dashed blue).

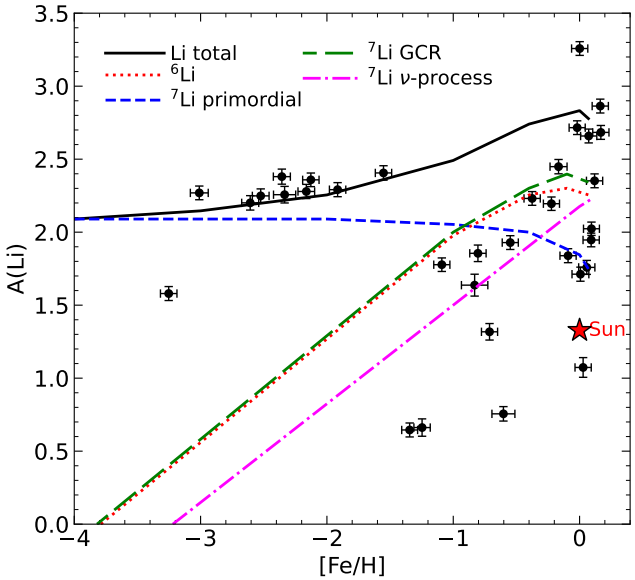

Figure 8. Lithium abundance as a function of [Fe/H]. The GCE models from B. D. Fields & K. A. Olive (1999a, 1999b) for different production channels are plotted, with the black line showing the total value. The solar fitted value is marked as a red star.

find a large scatter of A(Li) at higher metallicities ([Fe/H] $\gtrsim -1.5$). They noted that the poor fit in this [Fe/H] regime may the consequence of missing production sources of Li, such as AGBs and classical novae (D. Romano et al. 2021). This is because each star would have its own production history that cannot be captured by a unique GCE model. We note that the star with a very low Li abundance at [Fe/H] $= -3.25$ (HD 133442) is a cool giant ($T_{\rm eff} \approx 5520$ K), cooler than other metal-poor stars by ≈500 K. Therefore, the lower Li abundance is likely caused by destruction of it in the deeper convection envelope (see similar discussions on this in, e.g., F. Spite & M. Spite 1982).

4. Conclusions

We present our new NLTE astrophysical analysis code, which is based on the Payne ANN. This code will be used within the SAPP in the 4MOST Galactic Pipeline (4GP). Our code will derive stellar parameters and abundances in the 4MIDABLE-HR survey for over 4 million targets with the target SNR > 100 Å^{-1}. To construct the new ANN, we computed a new grid of 404,793 NLTE synthetic spectral models, broadly covering the entire parameter space of FGK-type stars and metallicities. Different architectures of the neural network, such as the activation functions, were carefully tested and adopted to give the lowest possible final training loss within the network requirements. We showed that SiLU activation functions halve the final loss, but also the ANN requires a sigmoid activation function in the output layer for best performance. The NLTE departure coefficients for the following 16 chemical elements are used: H, O, Na, Mg, Al,

Si, Ca, Ti, Mn, Fe, Co, Ni, Sr, Y, Ba, and Eu (see also J. M. Gerber et al. 2023).

We release the new NLTE astrophysical analysis code at our group website.[19] The online version of the code can be used to analyze 4MOST high-resolution stellar spectra in real time. The line masks used are provided in Table D1. The code used for training is publicly available as well.[20] Our NLTE ANN code enables a direct and self-consistent analysis of spectra and determination of stellar parameters $T_{\rm eff}$, $\log g$, [Fe/H], and $\xi_{\rm t}$; broadening $v_{\rm brd}$; and chemical abundances A(Li), [C/Fe], [O/Fe], [Na/Fe], [Mg/Fe], [Al/Fe], [Si/Fe], [Ca/Fe], [Ti/Fe], [Cr/Fe], [Mn/Fe], [Co/Fe], [Ni/Fe], [Sr/Fe], [Y/Fe], [Ba/Fe], and [Eu/Fe]. It takes 10–20 s for a single spectrum on a single laptop CPU core to derive all stellar parameters and abundances.

We also validated the NLTE ANN performance on 67 unique archival spectra of Gaia-ESO benchmark stars and 54 additional spectra of nearby bright stars. The sample comprises 84 stars, which cover the following parameter space: $3700 \lesssim T_{\rm eff}/{\rm K} \lesssim 6900$, $0.8 \lesssim \log g \lesssim 5.0$, and $-3.35 \lesssim$ [Fe/H] $\lesssim 0.34$. The observed spectra of these red giants, main-sequence stars, and subgiants were taken from HARPS, UVES, NARVAL, and FOCES and degraded to the 4MOST-HR resolution $R \approx 20{,}000$ to ensure consistency. For these stars, we also derive all chemical abundances using an independent code, TSFitPy (J. M. Gerber et al. 2023; N. Storm & M. Bergemann 2023), and we show that the NLTE Payne machine-learning-based abundances are in excellent agreement with the values obtained using classical spectroscopic methods.

We furthermore compare our new NLTE abundances with the GCE model OMEGA+ (B. Côté et al. 2017, 2018). In particular, we show how the SFE can be constrained using chemical elements such as Na, Al, O, Mg, Si, Sr, and Ba, favoring lower values of the SFE, $\epsilon_\star \lesssim 0.03$ for the Galactic disk. We show that the new NLTE Payne abundances demonstrate the same [X/Fe] behavior as shown in previous NLTE studies, including O (A. M. Amarsi et al. 2015), Na (G. Zhao et al. 2016), Mg (M. Bergemann et al. 2017), Al (G. Zhao et al. 2016), Si (G. Zhao et al. 2016), Ca (L. Mashonkina et al. 2017), Ti (M. Bergemann 2011), Mn (P. Eitner et al. 2020; N. Storm et al. 2025), Ni (P. Eitner et al. 2023; N. Storm et al. 2025), Sr, Y, Ba, and Eu (N. Storm et al. 2025). This work highlights the potential of 4MOST data in comprehensively constraining the Galactic chemical enrichment and the origins of chemical elements for the first time using the newly trained NLTE ANN.

Acknowledgments

We sincerely thank the referee for the positive feedback and comments that improved the readability of the paper.

Based on 4MOST observations collected at the European Southern Observatory under ESO program 116.29EL.001. Funding for the construction of 4MOST, its science operations, and data processing has been provided by the 4MOST Consortium: https://4most.eu/cons.

We thank the 4MOST IWG7 working group and Tadafumi Matsuno for valuable discussions. N.S., M.B., and G.G. acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 949173). M.B. is supported through the Lise Meitner grant from the Max Planck Society. We acknowledge support by the Collaborative Research center SFB 881 (projects A5, A10), Heidelberg University, of the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation). T.B. acknowledges financial support by grant No. 2024-04990 from the Swedish Research Council. G.G. acknowledges support by Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) project ID eBer-22-59652 (GU 2240/1-1 "Galactic Archaeology with Convolutional Neural-Networks: Realising the potential of Gaia and 4MOST"). Computations were performed on the HPC system Raven at the Max Planck Computing and Data Facility. V.K. acknowledges financial support from the German Excellence Strategy via the Heidelberg Cluster "STRUCTURES" (EXC 2181–390900948), the German Ministry for Economic Affairs and Climate Action in project "MAINN" (funding ID 50OO2206), and the Carl Zeiss Stiftung. G.K. acknowledges support from the Bourse Qualité Recherche of the Observatoire de la Côte d'Azur (BQR OCA). G.T. acknowledges the financial support of the Slovenian Research Agency (research core funding No. P1-0188) and the European Space Agency (Prodex Experiment Arrangement No. 4000143450).

Appendix A
Details on the Neural Network Training

In this section, we outline the technical details of our ANN training method. The Payne was implemented and trained using PyTorch (A. Paszke et al. 2019) and consists of three fully connected hidden layers of 1024 neurons each, using SiLU activation functions, and a final output layer of size 33,375 with a sigmoid activation function. The latter limits the output to within 0 to 1, consistent with the range of normalized spectra. The network size was chosen as a balance between complexity and the speed of the network (e.g., see Figure 4 in T. Różański et al. 2025). Figure A1 shows a summary of our optimization experiments that consisted of training a series of networks on a smaller training set of 288,211 spectra. In particular, we trained networks with two different configurations—one using rectified linear unit (ReLU) activation functions (black) and one that employed SiLU instead (red)—for a variety of initial learning rates between 0.0001 and 0.003. ReLU refers to a function $\mathrm{ReLU}(x) = \max(0, x)$, while SiLU is a smoother approximation, $\mathrm{SiLU}(x) = x \times \mathrm{sigmoid}(x)$. For SiLU, we also tested a network trained for half the steps (red cross) and using only half of the training spectra (red triangle). The main takeaway is that the right choice of activation function reduces the final loss significantly, in this case by around half. This is not unexpected, since SiLU is smoother than ReLU. This is the typical behavior we expect to see in stellar spectra—e.g., absorption lines become weaker and do not disappear in a step-function manner. We also did a brief test, skipping the final sigmoid layer after the output. This significantly increased the final loss by around a factor of 2. A too low or too high initial learning rate can also result in a suboptimal training convergence. For our network, reducing the number of training steps or training spectra by half had the smallest impact. Out of the final training set of 404,793 spectra, 6% were used for the validation set. The individual abundance values were chosen in a uniformly random fashion;

[19] https://nlte.mpia.de/gui-payne_fit.php
[20] https://github.com/stormnick/one_payne

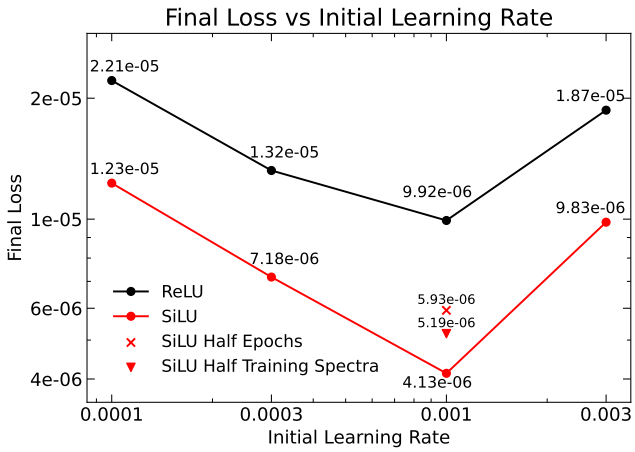

Figure A1. Final validation loss of the neural networks as a function of the initial learning rate. The neural networks indicated as the top black points were trained using ReLU activation functions, while the bottom red ones used SiLU instead. Except for the two separate activation functions and the initial learning rate, all of the networks were trained in an identical manner.

thus, the network a priori contains no elemental abundance bias and should be able to fit nonstandard chemical composition stars. Our training spectra did not contain any broadening. However, it was dynamically applied after the emulation from Payne during fitting, similar to how it is done with classical stellar synthesis codes. This simplifies the network, reduces the number of training spectra needed, and thus improves the convergence. The network was trained on four A100 GPUs. The loss was calculated using a mean-squared difference, AdamW optimizer with a cosine-type scheduler, which dynamically reduced the learning rate over time for better network convergence.

We did not highlight this in the figure, but the network size has a direct impact as well. We tried both smaller (256 neurons per hidden layer) and larger (2048 neurons per hidden layer) networks. In our case, a larger network did not significantly decrease the final loss—by only around 20%. However, a smaller network had a bigger impact—a 4.2 times smaller network had a 3.8 times larger loss (256 neurons per hidden layer resulted in a loss of 1.56×10^{-5}). We also tested the performance of the smaller network on the stellar parameters and abundances. The stellar parameters and most of the abundances had a nearly identical recovery rate as the big network, with only a slight increase in bias and spread compared to TSFitPy values (typically by around 0.02 dex). However, both Y and Eu bias suffered significantly, especially at the low abundance ranges. For europium, the average bias went to 0.3 dex (up to 0.7–1.0 dex for the lowest abundances). While the effect of the larger loss was not seen for most elements, it is more noticeable for weaker lines. Thus, we decided to stick with the network that has 1024 neurons per hidden layer as the balance between inference time and accuracy of the network.

The final technical detail that is relevant for this network architecture is the potential to decrease inference time. Let us imagine that one needs to fit only one specific line in the spectrum. In this case, skipping the synthesis of the remaining parts of the spectrum could speed up the process. This is possible for our Payne network. Since the network is constructed as an input layer connected with hidden layers and the final output layer, one is only required to do all of the initial matrix multiplications first. The calculation in the output layer can skip all of the pixels except for the ones that are of interest. The last matrix is the largest one by far ($1024 \times 33{,}375$ in our case, compared to 1024×1024 and 21×1024 for the other ones), so skipping calculations in the final matrix can give a significant time gain, despite using a "big" Payne network.

Appendix B
Validation of the Network Performance

We also tested the performance of the network across the parameter space. Figure B1 shows the mean rms normalized flux error for the validation synthetic spectra set across the full parameter space. The RMSE was calculated by using input parameters for each synthetic spectrum from the validation set, producing a Payne-equivalent spectrum and calculating an error compared to the original input Turbospectrum model using

$$\mathrm{RMSE} = \sqrt{\frac{1}{P}\sum_{j=1}^{P}\left(\hat{y}_j - y_j\right)^2}, \tag{B1}$$

where P is the number of pixels, $\hat{y}_j$ is the predicted spectrum from Payne, y_j is the input spectrum from Turbospectrum, and j is each pixel. On average, the blue window shows the largest error, due to the dominance of molecular bands. A large error is present in the cores of the CH G band at the ≈4300 Å lines but also in the cores of the H I and Li I lines. On average, the error is <0.005 (less than 0.5%) in normalized flux units, which implies that the error associated with the Payne interpolation is also typically <0.005 flux units.

Figure B2 shows the mean rms flux error across each spectrum for the full parameter space range. Based on the $T_{\rm eff}$–$\log g$ diagram, lower $T_{\rm eff}$ values result in higher RMSE because of the stronger molecular lines. Thus, the spectra with $T_{\rm eff} > 5000$ K have a mean RMSE < 0.002, also implying that lower [Fe/H] spectra at the same $T_{\rm eff}$ have a smaller RMSE. Typically, individual abundance values do not have an effect on the RMSE, except for the abundance of carbon. [C/Fe] ≈ 2 at the lowest $T_{\rm eff}$ results in the highest RMSE values.

In Figure B3, we show the recovery rate of true abundances for a fraction of synthetic spectra in the validation set with added noise (SNR = 100 and 250 $\mathrm{Å}^{-1}$). Black points indicate that the abundance was recovered in spectra with at least SNR = 100 $\mathrm{Å}^{-1}$, red points were recovered only for SNR = 250 $\mathrm{Å}^{-1}$, and gray ones indicate spectra with nonrecovered abundance. This shows the best possible performance limits of the current fitting with this network, although we note that a different line mask selection or fitting algorithm would change the distribution. Overall, the results are encouraging: metallicity (parameterized by [Fe/H]) is recovered in almost all spectra, except for the lowest-metallicity regime in the hottest stars at lower SNR. Each element group shows a good recovery rate across all metallicity regimes: Mg for α-elements, Mn for iron-peak elements, and Sr for neutron-capture elements. Li recovery depends directly on the stellar effective temperature, while O and Al have the worst recovery properties due to the weakness of their spectral lines.

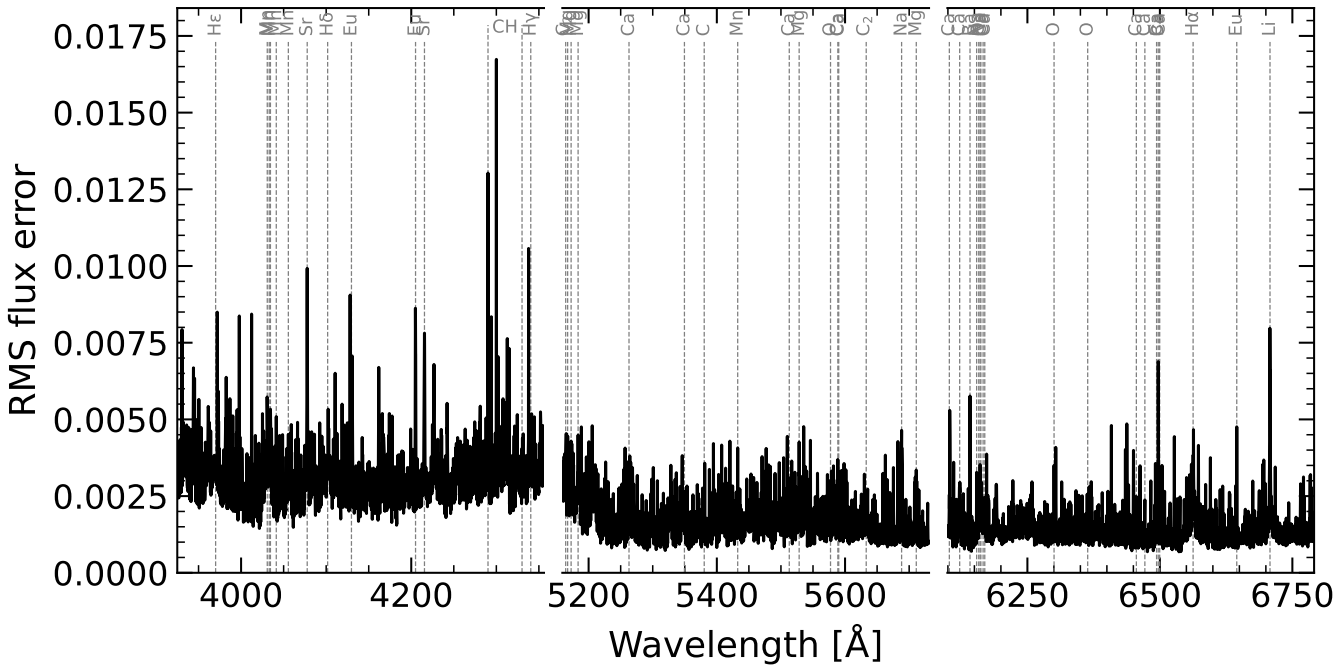

Figure B1. Mean rms normalized flux error across the stellar spectrum range for all validation synthetic spectra. Some of the notables lines are overplotted as well.

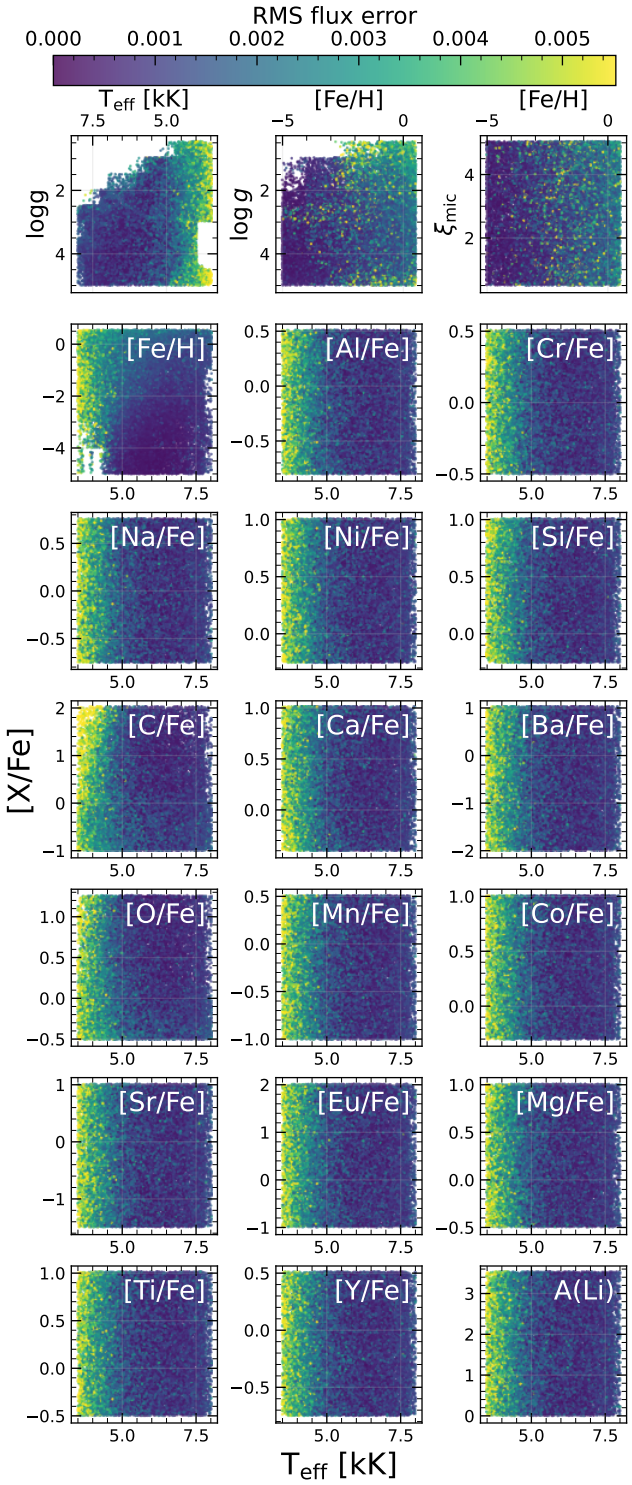

Figure B2. Distribution of the mean rms flux error of each spectrum (averaged across wavelength) in a 2D histogram of the validation set of synthetic spectra. The top left panel shows the $T_{\rm eff}$ diagram, the other two top panels are distributions as a function of [Fe/H], and the rest are as a function of $T_{\rm eff}$. This distribution is shown because $T_{\rm eff}$ has the largest impact on the RMSE compared to any parameter in the grid.

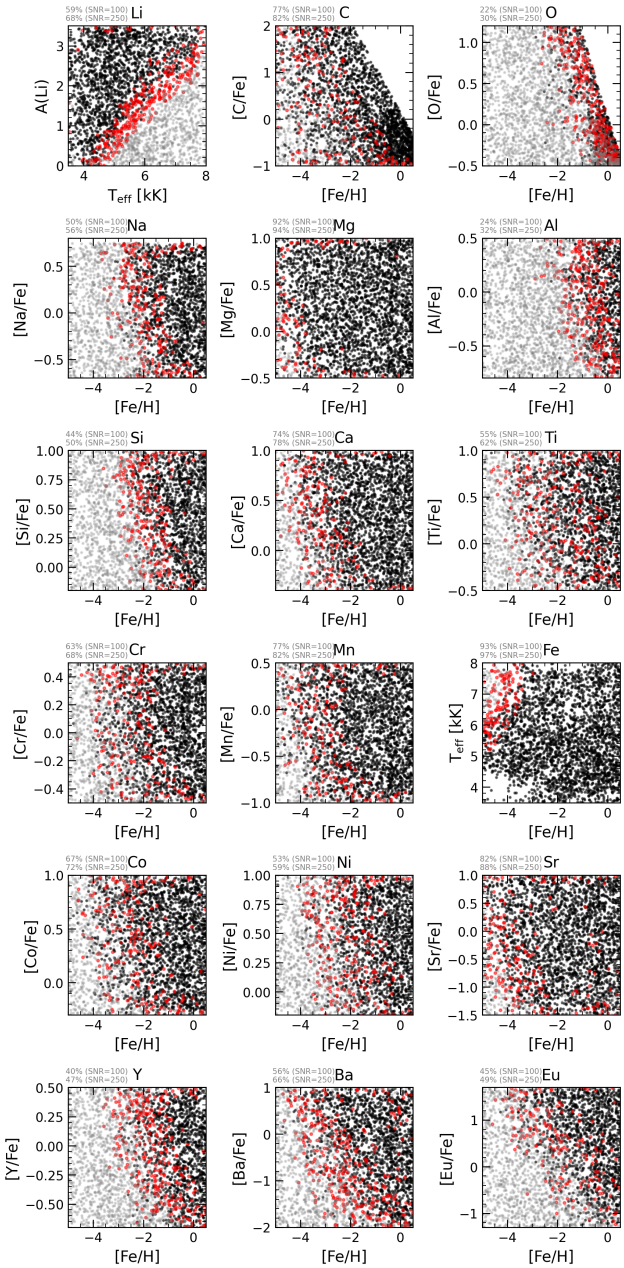

Figure B3. Recovery rate of the true abundances for the fraction of the validation set of synthetic spectra with added noise (SNR = 100 and 250 Å^{-1}). Black points indicate abundances recovered at least for SNR = 100 Å^{-1}, red points indicate abundances recovered for SNR = 250 Å^{-1}, and gray points indicate nonrecovered abundances. The number in the top left of each subplot indicates the total recovery percentage.

Appendix C
Results for the Solar Spectra

In Table C1, we present the average abundances obtained using five high-resolution solar spectra (NARVAL, UVES, and HARPS spectrographs), which we use to self-consistently compute the

Table C1
Average Fitted Solar Abundances Based on Five 4MOSTified Solar Spectra

Element	Fitted Abundance
A(Al)	6.45
A(Ba)	2.15
A(C)	8.43
A(Ca)	6.43
A(Co)	4.96
A(Cr)	5.65
A(Eu)	0.42
A(Fe)	7.45
A(Mg)	7.53
A(Mn)	5.44
A(Na)	6.29
A(Ni)	6.31
A(O)	8.74
A(Si)	7.50
A(Ti)	5.08
A(Y)	1.98

stellar abundances in the square bracket notation. The solar abundance of Y is not very accurate, possibly either because of the use of older sources of $\log gf$ values (see also discussion in Section 2.2. in N. Storm et al. 2024) or due to the lack of strong reliable diagnostic lines in the 4MOST-HR spectral range.

Appendix D
Lines Used for Fitting

In Table D1 we provide the list of lines for each individual element that were used to fit stellar abundances with the corresponding atomic data.

Table D1
Atomic Data of the Lines Used in the Fitting of Different Elements

Element	λ (Å)	$E_{\rm low}$ (eV)	$\log(gf)$	$g_{\rm up}$	Isotope (if any)
Al I	6696.023	3.143	−1.569	4	...
Al I	6698.673	3.143	−1.870	2	...
Ba II	6141.709	0.704	−0.395	4	137
Ba II	6141.713	0.704	−0.032	4	134
Ba II	6141.714	0.704	−0.032	4	136

(This table is available in machine-readable form in the online article.)

Appendix E
Fitted Parameters for All Spectra

In Table E1 we provide fitted stellar parameters for all individual spectra before averaging their values and correcting them for the solar abundances from Appendix C.

Table E1
Fitted Parameters for All Spectra

Star	$T_{\rm eff}$ (K)	$\log g$	$\xi_{\rm t}$ (km s^{-1})	[Fe/H]	$v_{\rm brd}$ (km s^{-1})	[Al/Fe]	[Cr/Fe]	[Na/Fe]
15PEG[5]	6488 ± 137	4.11 ± 0.25	1.58 ± 0.01	−0.47 ± 0.06	11.18 ± 0.03	⋯	−0.16 ± 0.13	0.10 ± 0.08
18Sco[1]	5828 ± 135	4.49 ± 0.25	1.36 ± 0.01	−0.05 ± 0.06	1.39 ± 0.02	0.12 ± 0.07	−0.06 ± 0.11	0.02 ± 0.08
18Sco[2]	5835 ± 134	4.52 ± 0.24	1.23 ± 0.00	−0.04 ± 0.06	1.41 ± 0.01	0.10 ± 0.07	−0.05 ± 0.10	0.01 ± 0.07
31AQL[5]	5575 ± 136	4.23 ± 0.25	1.10 ± 0.01	0.34 ± 0.06	4.26 ± 0.04	0.12 ± 0.08	−0.01 ± 0.18	0.19 ± 0.08
Arcturus[2]	4165 ± 134	1.30 ± 0.25	1.58 ± 0.00	−0.72 ± 0.06	2.84 ± 0.02	0.35 ± 0.07	0.05 ± 0.11	0.14 ± 0.07

Note. The uncertainty includes both systematic and errors returned by `LMFIT`. Instruments: (1) UVES, (2) NARVAL, (3) UVES-POP, (4) HARPS, (5) FOCES. (This table is available in machine-readable form in the online article.)

ORCID iDs

Nicholas Storm https://orcid.org/0000-0002-5259-3974
Maria Bergemann https://orcid.org/0000-0002-9908-5571
Tomasz Różański https://orcid.org/0000-0002-5819-3023
Victor F. Ksoll https://orcid.org/0000-0002-0294-799X
Thomas Bensby https://orcid.org/0000-0003-3978-1409
Gregor Traven https://orcid.org/0000-0002-1391-9097
Georges Kordopatis https://orcid.org/0000-0002-9035-3920
Ross P. Church https://orcid.org/0000-0001-9204-0779
Weijia Sun https://orcid.org/0000-0002-3279-0233
Gražina Tautvaišienė https://orcid.org/0000-0001-7672-154X

References

Amarsi, A. M., Asplund, M., Collet, R., & Leenaarts, J. 2015, MNRAS, 454, L11
Arcones, A., & Thielemann, F.-K. 2023, A&ARv, 31, 1
Arnould, M., Goriely, S., & Takahashi, K. 2007, PhR, 450, 97
Asplund, M. 2005, ARA&A, 43, 481
Barbuy, B., Chiappini, C., & Gerhard, O. 2018, ARA&A, 56, 223
Battistini, C., & Bensby, T. 2015, A&A, 577, A9
Battistini, C., & Bensby, T. 2016, A&A, 586, A49
Beniamini, P., & Piran, T. 2019, MNRAS, 487, 4847
Bensby, T., Bergemann, M., Rybizki, J., et al. 2019, Msngr, 175, 35
Bensby, T., Feltzing, S., & Oey, M. S. 2014, A&A, 562, A71
Bergemann, M. 2011, MNRAS, 413, 2184
Bergemann, M., Collet, R., Amarsi, A. M., et al. 2017, ApJ, 847, 15
Bergemann, M., Gallagher, A. J., Eitner, P., et al. 2019, A&A, 631, A80
Bergemann, M., & Gehren, T. 2008, A&A, 492, 823
Bergemann, M., Hansen, C. J., Bautista, M., & Ruchti, G. 2012a, A&A, 546, A90
Bergemann, M., & Hoppe, R. 2025, arXiv:2511.04254
Bergemann, M., Hoppe, R., Semenova, E., et al. 2021, MNRAS, 508, 2236
Bergemann, M., Kudritzki, R.-P., Würl, M., et al. 2013, ApJ, 764, 115
Bergemann, M., Lind, K., Collet, R., Magic, Z., & Asplund, M. 2012b, MNRAS, 427, 27
Bergemann, M., Pickering, J. C., & Gehren, T. 2010, MNRAS, 401, 1334
Blanco-Cuaresma, S., Soubiran, C., Jofré, P., & Heiter, U. 2014, A&A, 566, A98
Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Buder, S., Kos, J., Wang, X. E., et al. 2025, PASA, 42, e051
Conroy, C. 2013, ARA&A, 51, 393
Côté, B., O'Shea, B. W., Ritter, C., Herwig, F., & Venn, K. A. 2017, ApJ, 835, 128
Côté, B., Silvia, D. W., O'Shea, B. W., Smith, B., & Wise, J. H. 2018, ApJ, 859, 67
Cowan, J. J., Sneden, C., Lawler, J. E., et al. 2021, RvMP, 93, 015002
Cristallo, S., Straniero, O., Piersanti, L., & Gobrecht, D. 2015, ApJS, 219, 40
de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, Msngr, 175, 3
Eitner, P., Bergemann, M., Hansen, C. J., et al. 2020, A&A, 635, A38
Eitner, P., Bergemann, M., Ruiter, A. J., et al. 2023, A&A, 677, A151
Ezzeddine, R., Merle, T., Plez, B., et al. 2018, A&A, 618, A141
Fields, B. D., & Olive, K. A. 1999a, ApJ, 516, 797
Fields, B. D., & Olive, K. A. 1999b, NewA, 4, 255
Forestini, M., & Charbonnel, C. 1997, A&AS, 123, 241
Fuhrmann, K. 1998, A&A, 338, 161
Fuhrmann, K., Axer, M., & Gehren, T. 1993, A&A, 271, 451
Gallagher, A. J., Bergemann, M., Collet, R., et al. 2020, A&A, 634, A55
Gehren, T., Liang, Y. C., Shi, J. R., Zhang, H. W., & Zhao, G. 2004, A&A, 413, 1045
Gehren, T., Shi, J. R., Zhang, H. W., Zhao, G., & Korn, A. J. 2006, A&A, 451, 1065
Gent, M. R., Bergemann, M., Serenelli, A., et al. 2022, A&A, 658, A147
Gerber, J. M., Magg, E., Plez, B., et al. 2023, A&A, 669, A43
Gilmore, G., Wyse, R. F. G., & Jones, J. B. 1995, AJ, 109, 1095
Grevesse, N., Asplund, M., & Sauval, A. J. 2007, SSRv, 130, 105
Grevesse, N., & Sauval, A. J. 1998, SSRv, 85, 161
Guiglion, G., Nepal, S., Chiappini, C., et al. 2024, A&A, 682, A9
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hegedüs, V., Mészáros, S., Világos, B., et al. 2025, A&A, 699, A293
Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, A&A, 582, A49
Heiter, U., Lind, K., Bergemann, M., et al. 2021, A&A, 645, A106
Karakas, A. I., & Lattanzio, J. C. 2003, PASA, 20, 279
Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, ApJ, 900, 179
Kordopatis, G., Hill, V., & Lind, K. 2023, A&A, 674, A104
Kovalev, M., Bergemann, M., Ting, Y.-S., & Rix, H.-W. 2019, A&A, 628, A54
Lawler, J. E., Wickliffe, M. E., den Hartog, E. A., & Sneden, C. 2001, ApJ, 563, 1075
Lian, J., Storm, N., Guiglion, G., et al. 2023, MNRAS, 525, 1329
Limongi, M., & Chieffi, A. 2018, ApJS, 237, 13
Lind, K., & Amarsi, A. M. 2024, ARA&A, 62, 475
Magg, E., Bergemann, M., Serenelli, A., et al. 2022, A&A, 661, A140
Maiolino, R., & Mannucci, F. 2019, A&ARv, 27, 3
Mashonkina, L., Jablonka, P., Sitnova, T., Pakhomov, Y., & North, P. 2017, A&A, 608, A89
Mashonkina, L., Zhao, G., Gehren, T., et al. 2008, A&A, 478, 529
Matteucci, F. 2021, A&ARv, 29, 5
Mishenina, T., Pignatari, M., Côté, B., et al. 2017, MNRAS, 469, 4378
Mowlavi, N. 1999, A&A, 350, 73
Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16
Newville, M., Otten, R., Nelson, A., et al. 2025, LMFIT: Non-Linear Least-Squares Minimization and Curve-Fitting for Python, v1.3.4, Zenodo, doi:10.5281/zenodo.16175987
Nishimura, N., Takiwaki, T., & Thielemann, F.-K. 2015, ApJ, 810, 109
Nordlund, Å., Stein, R. F., & Asplund, M. 2009, LRSP, 6, 2
Pancino, E., Lardo, C., Altavilla, G., et al. 2017, A&A, 598, A5
Paszke, A., Gross, S., Massa, F., et al. 2019, arXiv:1912.01703
Prantzos, N., Abia, C., Limongi, M., Chieffi, A., & Cristallo, S. 2018, MNRAS, 476, 3432
Reetz, J. K. 1991, diploma thesis, Universität München
Romano, D., Magrini, L., Randich, S., et al. 2021, A&A, 653, A72
Różański, T., & Ting, Y.-S. 2025, OJAp, 8, 69
Różański, T., Ting, Y.-S., & Jabłońska, M. 2025, ApJ, 980, 66
Ryabchikova, T., Piskunov, N., Kurucz, R. L., et al. 2015, PhyS, 90, 054005
Ryan, S. G., Beers, T. C., Olive, K. A., Fields, B. D., & Norris, J. E. 2000, ApJL, 530, L57
Schönrich, R., & Bergemann, M. 2014, MNRAS, 443, 698
Semenova, E., Bergemann, M., Deal, M., et al. 2020, A&A, 643, A164
Serenelli, A. M., Bergemann, M., Ruchti, G., & Casagrande, L. 2013, MNRAS, 429, 3645
Sneden, C. A. 1973, PhD thesis, Univ. Texas
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
Soubiran, C., Creevey, O. L., Lagarde, N., et al. 2024, A&A, 682, A145
Spite, F., & Spite, M. 1982, A&A, 115, 357
Storm, N., Barklem, P. S., Yakovleva, S. A., et al. 2024, A&A, 683, A200
Storm, N., & Bergemann, M. 2023, MNRAS, 525, 3718
Storm, N., Bergemann, M., Eitner, P., et al. 2025, MNRAS, 538, 3284
Ting, Y. S., Conroy, C., Rix, H. W., & Cargile, P. 2019, ApJ, 879, 69
Tolstoy, E., Hill, V., & Tosi, M. 2009, ARA&A, 47, 371
Venn, K. A., Irwin, M., Shetrone, M. D., et al. 2004, AJ, 128, 1177
Voronov, Y. V., Yakovleva, S. A., & Belyaev, A. K. 2022, ApJ, 926, 173
Walcher, C. J., Banerji, M., Battistini, C., et al. 2019, Msngr, 175, 12
Wedemeyer, S., Kučinskas, A., Klevas, J., & Ludwig, H.-G. 2017, A&A, 606, A26
Wehrhahn, A., Piskunov, N., & Ryabchikova, T. 2023, A&A, 671, A171
Wheeler, A. J., Abruzzo, M. W., Casey, A. R., & Ness, M. K. 2023, AJ, 165, 68
Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, ApJS, 245, 34
Yakovleva, S. A., Belyaev, A. K., & Bergemann, M. 2020, Atoms, 8, 34
Zhang, M., Xiang, M., Ting, Y.-S., et al. 2025, ApJS, 279, 5
Zhao, G., Mashonkina, L., Yan, H. L., et al. 2016, ApJ, 833, 225