

# 3D NLTE spectral line formation of lithium in late-type stars

Ella Xi Wang<sup>1</sup>,<sup>1,2</sup>★ Thomas Nordlander<sup>1,2</sup>,<sup>1,2</sup> Martin Asplund,<sup>3</sup> Anish M. Amarsi<sup>1,4</sup>,<sup>4</sup> Karin Lind<sup>5</sup> and Yixiao Zhou<sup>1</sup>

<sup>1</sup>Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

<sup>2</sup>ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Canberra, ACT 2601, Australia

<sup>3</sup>Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85741 Garching, Germany

<sup>4</sup>Theoretical Astrophysics, Department of Physics and Astronomy, Uppsala University, Box 516, SE-751 20 Uppsala, Sweden

<sup>5</sup>Department of Astronomy, Stockholm University, AlbaNova University Center, SE-106 91 Stockholm, Sweden

Accepted 2020 October 23. Received 2020 October 22; in original form 2020 July 29

## ABSTRACT

Accurately known stellar lithium abundances may be used to shed light on a variety of astrophysical phenomena such as big bang nucleosynthesis, radial migration, ages of stars and stellar clusters, and planet engulfment events. We present a grid of synthetic lithium spectra that are computed in non-local thermodynamic equilibrium (NLTE) across the STAGGER grid of three-dimensional (3D) hydrodynamic stellar atmosphere models. This grid covers three Li lines at 610.4, 670.8, and 812.6 nm for stellar parameters representative of FGK-type dwarfs and giants, spanning  $T_{\text{eff}} = 4000\text{--}7000$  K,  $\log g = 1.5\text{--}5.0$ ,  $[\text{Fe}/\text{H}] = -4.0\text{--}0.5$ , and  $A(\text{Li}) = -0.5\text{--}4.0$ . We find that our abundance corrections are up to 0.15 dex more negative than in previous work, due to a previously overlooked NLTE effect of blocking of UV lithium lines by background opacities, which has important implications for a wide range of science cases. We derive a new 3D NLTE solar abundance of  $A(\text{Li}) = 0.96 \pm 0.05$ , which is 0.09 dex lower than the commonly used value. We make our grids of synthetic spectra and abundance corrections publicly available through the BREIDABLIK package. This package includes methods for accurately interpolating our grid to arbitrary stellar parameters through methods based on Kriging (Gaussian process regression) for line profiles, and multilayer perceptrons (a class of fully connected feedforward neural networks) for NLTE corrections and 3D NLTE abundances from equivalent widths, achieving interpolation errors of the order of 0.01 dex.

**Key words:** line: profiles – radiative transfer – stars: abundances – stars: atmospheres – stars: late-type.

## 1 INTRODUCTION

Lithium is the only element which can be produced through big bang nucleosynthesis (BBN), cosmic-ray spallation, and stellar processes. Lithium is also a fragile element, destroyed in low-mass stars in a manner that depends on stellar mass, metallicity, age, and possibly other factors. The many production and destruction processes make lithium a complex but insightful element. Lithium can thus illuminate a number of distinct science cases, including the cosmological lithium problem, radial migration, ages of stars and stellar clusters, and planet engulfment events. However, these applications in large part rely on accurate measurements of lithium abundances in late-type stars.

The lithium abundance ( $A(\text{Li})^1$ ) in stars can be used to estimate the cosmological value produced during BBN (Cyburt et al. 2016). Warm old, metal-poor dwarfs near the main-sequence turnoff-off (MSTO) have been found to exhibit roughly the same lithium abundances over a wide range of metallicities, the so-called Spite plateau (Spite

& Spite 1982). This plateau has  $A(\text{Li}) \approx 2.0\text{--}2.2$ , and is commonly interpreted to reflect the lithium abundance with which these old metal-poor stars were born. However, the cosmological lithium abundance predicted from BBN is  $A(\text{Li}) = 2.75 \pm 0.02$ , a factor of at least three larger than the abundance in the oldest stars (Pitrou et al. 2018). The reason for this discrepancy is generally thought to be due to a combination of depletion and non-destructive deposition through gravitational settling (e.g. Richard, Michaud & Richer 2005) but speculation also exists that this may signal non-standard particle physics affecting the BBN production (Fields 2011). Parameterized stellar Li depletion predictions reproduce observations of lithium in globular cluster stars quite well, including the signature of dredge-up through a short-lived increase of lithium at the middle of the subgiant branch (e.g. Korn et al. 2007; Lind et al. 2008; Nordlander et al. 2012; Gruyters, Nordlander & Korn 2014; Gruyters et al. 2016). Mucciarelli, Salaris & Bonifacio (2012) recognized that lithium abundances in metal poor lower RGB stars also exhibit a plateau. These abundances are sensitive only to effects of dilution (reflecting the depth of the convection zone) and destruction that occurred prior, while effects of non-destructive deposition are erased by the deep convection zone. Measurements for RGB stars in the Milky Way,  $A(\text{Li}) \approx 1.0$  (Mucciarelli et al. 2012), as well as in the extragalactic globular cluster M54,  $A(\text{Li}) \approx 0.9$  (Mucciarelli et al. 2014), imply an initial composition with  $A(\text{Li}) \approx 2.3\text{--}2.4$  that does

\* E-mail: [ellawang@mso.anu.edu.au](mailto:ellawang@mso.anu.edu.au)

<sup>1</sup>We use the customary abundance notation where  $A(X) \equiv \log(N_X/N\text{H}) + 12$  and  $[X/Y] \equiv (A(X) - A(Y)) - (A(X) - A(Y))_{\odot}$ , with  $N_X$  representing the number density of element “X”.

not vary with environment and therefore is unlikely to be depleted through external mechanisms. Although slightly higher than the Spite plateau, these estimates still imply significant destruction relative to the cosmological abundance.

At metallicities  $[\text{Fe}/\text{H}] > -1$ , the abundance of lithium becomes coupled to  $[\text{Fe}/\text{H}]$  through chemical enrichment processes such as cosmic-ray spallation (Prantzos 2012; Prantzos et al. 2017) and possibly also stellar production (Bensby & Lind 2018). Lithium abundance measurements in stars born at different locations and times can therefore tell us about the history of chemical evolution in the Galaxy, and radial migration of stars in the Galaxy (Schönrich & Binney 2009; Prantzos et al. 2017).

Lithium abundance measurements can be used to determine ages for young stellar clusters. Because pre-main-sequence stars are fully convective, they will rapidly deplete lithium when their cores reach  $T \approx 2.5 \times 10^6$  K (Basri, Marcy & Graham 1996; Chabrier, Baraffe & Plez 1996; Bildsten et al. 1997; Ushomirsky et al. 1998; Jeffries 2014). Even in coeval groups, the amount of depletion does, however, vary significantly from star to star (e.g. Sestito & Randich 2005; Žerjal et al. 2019), possibly tied to details of rotation and angular momentum transport (see e.g. Do Nascimento et al. 2009; Soderblom 2010). It is therefore still unclear just why stars like the Sun exhibit significant lithium depletion by more than two orders of magnitude (e.g. Meléndez et al. 2010). Similar patterns are seen in other stars, where it is still unclear whether this depletion correlates with the presence of exoplanets (Gonzalez 2008; Israelian et al. 2009; Delgado Mena et al. 2015; Gonzalez 2015) as several studies find no such correlation (Ryan 2000; Baumann et al. 2010; Ramírez et al. 2012; Bensby & Lind 2018; Carlos et al. 2019). Some of the proposed mechanisms for enhanced lithium depletion for stars with planets are stellar angular momentum loss due to planetary migration (Castro et al. 2008), and strong differential rotation caused by interactions between the protoplanetary disc and star (Bouvier 2008). Where no enhanced lithium depletion is detected, lithium depletion is instead linked to stellar ages; with older stars exhibiting increased lithium depletion (Baumann et al. 2010; Carlos et al. 2019).

In one-dimensional (1D) hydrostatic atmospheres, the accuracy of abundance determinations in late-type stars may be severely impacted by the commonly used approximations of local thermodynamic equilibrium (LTE). Modelling of lithium spectra in non-LTE (NLTE) has already been performed in many studies across a wide parameter space (e.g. Steenbock & Holweger 1984; Carlsson et al. 1994; Pavlenko & Magazzu 1996; Takeda & Kawanomoto 2005; Lind, Asplund & Barklem 2009; Osorio et al. 2011; Takeda 2019). These studies find that the NLTE effect can lead to abundance corrections of up to 0.4 dex (Lind et al. 2009).

In addition to LTE assumptions, 3D hydrodynamic atmospheres also affect measured lithium abundances. Recent studies suggest that although the differences between 3D NLTE and 1D LTE synthetic spectra have been found to often be rather small due to fortuitous cancellation between NLTE and 3D hydrodynamic effects that work in opposite directions, this cancellation is not perfect and abundance corrections may still be as large as 0.3 dex (Klevas et al. 2018). Full 3D NLTE calculations (e.g. Kiselman 1997; Asplund, Carlsson & Botnen 2003; Barklem, Belyaev & Asplund 2003; Sbordone et al. 2010; Lind et al. 2013; Klevas et al. 2016; Mott et al. 2017; Harutyunyan et al. 2018; Klevas et al. 2018; Mott et al. 2020) have so far been limited in their coverage of parameter space due to the large computational cost involved.

This study presents a 3D NLTE Li grid spanning the full parameter range expected for FGK-type dwarfs and giants, covering a subset of  $T_{\text{eff}} = 4000\text{--}7000$  K,  $\log g = 1.5\text{--}5.0$ ,  $[\text{Fe}/\text{H}] = -4\text{--}0.5$  – a total

of 195 3D hydrodynamic model atmospheres – with abundances in the range  $A(\text{Li}) = -0.5\text{--}4$ . We detail the model atom and stellar atmospheres in Section 2. In Section 3, we present and discuss our results with comparison to previous work. We introduce our interpolation methods and the publicly available synthetic spectra, spectrum interpolation, and abundance correction package BREIDABLIK<sup>2</sup> in Section 4. Lastly, in Section 5, we present results of our new lower inferred abundances on a handful of science cases, before presenting our conclusions in Section 6.

## 2 3D NLTE SPECTRAL LINE FORMATION FOR LI

We use the 3D NLTE radiative transfer code BALDER, which originates in the MULTI3D code (Botnen & Carlsson 1999; Leenaarts & Carlsson 2009) but has since been developed significantly, including a new equation-of-state and opacity package BLUE (Amarsi et al. 2016a,b, 2018).

The statistical equilibrium is solved by calculating the mean radiation field with short characteristic rays. These are distributed as  $\mu \equiv \cos \theta$  relative to the vertical on the interval  $-1$  to  $+1$  using a Gauss–Lobatto quadrature with eight points. For non-vertical rays, we consider four equidistant azimuthal angles, giving a total of 26 rays. The emergent flux is computed by calculating the emergent intensity in the vertical direction, as well as in seven inclined directions each using eight horizontal angles, for a total of 57 rays. We compute spectra for abundances in the range  $A(\text{Li}) = -0.5$  to  $+4.0$ , in steps of 0.5 dex, which produces an abundance interpolation error of at most  $\sim 0.02$  dex, as determined through comparison to calculations over the full abundance range with very small steps.

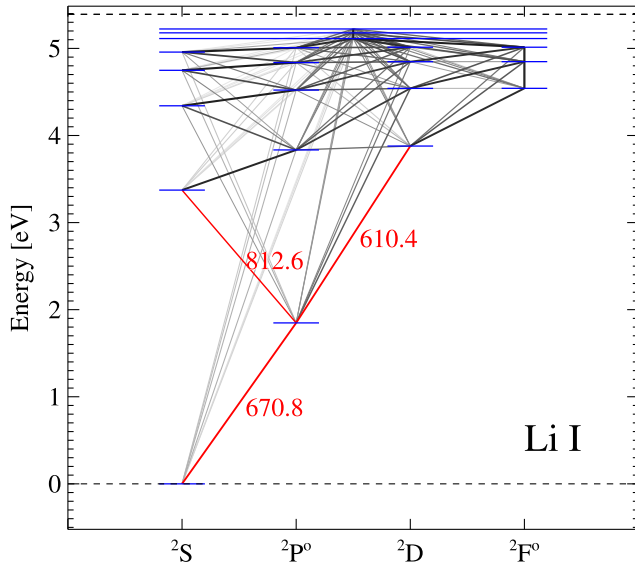
### 2.1 Li model atom

We use a model atom containing 20 levels of Li I plus the Li II ground state, connected by 113 bound–bound and 20 bound-free transitions. We compute the spectra of the three strongest transitions, highlighted in Fig. 1.

The model atom originates with Carlsson et al. (1994), and was substantially updated by Lind et al. (2009) and Osorio et al. (2011). Briefly, energy levels and radiative transition data come from TOPbase (Peach, Saraph & Seaton 1988), with some notable exceptions. The atomic data for the 670.8-nm transition uses an oscillator strength from Yan, Tambasco & Drake (1998), fine-structure splitting from Sansonetti et al. (2011) and hyperfine splitting from Beckmann, Böklen & Elke (1974) and Puchalski & Pachucki (2009). Although updated, the resulting wavelengths and relative oscillator strengths are consistent with those of Smith, Lambert & Nissen (1998). Likewise, the 610.4-nm transition uses an oscillator strength from Yan et al. (1998) and fine-structure splitting from Lindgård & Nielson (1977). For all transitions, we assume the presence of <sup>7</sup>Li only, neglecting any isotopic splitting. The three strongest transitions use collisional broadening parameters based on Anstee & O’Mara (1995) and Barklem & O’Mara (1997).

Cross-sections for excitation and ionization through electron collisions were calculated and implemented by Osorio et al. (2011). Inelastic hydrogen collisional transition rates for excitation and charge transfer of low-lying states (Barklem et al. 2003) are based on cross-sections from Croft, Dickinson & Gadea (1999) and Belyaev & Barklem (2003). We have further implemented inelastic hydrogen

<sup>2</sup><https://github.com/ellawang44/Breidablik>



**Figure 1.** Term diagram of lithium, illustrating the energies of different states, ordered according to spectroscopic term. Bound-bound radiative transitions are marked with black lines, and the abundance diagnostic lines at 610.4, 670.8, and 812.6 nm are highlighted in red. The long horizontal blue lines are super levels where the  $l$  quantum number has been collapsed. The  ${}^2F^{\circ}$  states are superstates representing  $l \geq 3$ .

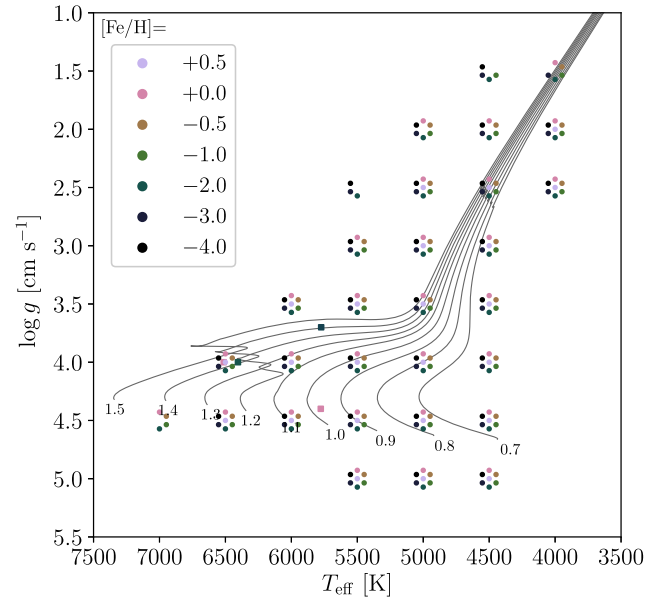
collisional excitation rates for more highly excited states following Kaulakys (1985, 1991), using the publicly available code KAULAKYS (Barklem 2016).

## 2.2 Stellar atmospheres

### 2.2.1 3D hydrodynamical stellar model atmospheres

To calculate 3D NLTE synthetic spectra, we use BALDER to perform radiative transfer post-processing of 3D hydrodynamical model atmospheres from the STAGGER-grid (Magic et al. 2013). Each simulation covers a time sequence of roughly two convective turnover times, represented by about 150 snapshots. A small number of simulations suffered from convergence problems, and are therefore not considered in this work. As a result, we selected 195 models, shown in Fig. 2. These cover stellar parameters representing FGK-type dwarfs and giants, in a wide range of metallicities between  $[\text{Fe}/\text{H}] = -4.0$  and  $+0.5$ . All models adopt the Asplund et al. (2009) metal mixture, aside from models with  $[\text{Fe}/\text{H}] \leq -1$  where an  $\alpha$ -enhancement of  $[\alpha/\text{Fe}] = 0.4$  dex is further applied.

From each simulation, we select five temporally equidistant snapshots. Test calculations based on larger sets of snapshots indicate that using five snapshots results in a typical abundance error ( $\sigma/\sqrt{N}$ ) less than 0.02 dex and 0.01 dex in LTE and NLTE respectively based on temporal variations in the equivalent widths. More specifically, we find the largest snapshot-to-snapshot variations amongst the hottest models, with equivalent widths varying by typically  $\sigma = 0.03$  dex in LTE and by 0.02 dex in NLTE. Importantly, the ratio of LTE and NLTE equivalent widths may have a larger standard deviation, of the order of 0.04 dex, indicating that temporal variations may have different sign in LTE and NLTE and do not perfectly correlate. This thus indicates that fewer snapshots are required to sample a simulation with NLTE calculations than the corresponding LTE calculations. These test calculations furthermore suggest that one should avoid the approach where NLTE corrections are derived from



**Figure 2.** Stellar parameters of the Stagger-grid models (Magic et al. 2013) used in this work, colour-coded according to  $[\text{Fe}/\text{H}]$ ; the Sun, HD 84937, HD 140283 and Procyon are shown with squares. Stellar evolution tracks at solar metallicity with masses in the range  $0.7\text{--}1.5 M_{\odot}$  in steps of  $0.1 M_{\odot}$  from MIST (Paxton et al. 2011, 2013, 2015; Choi et al. 2016; Dotter 2016) are shown for reference, labelled according to mass.

a small number of snapshots and applied to a longer sequence of LTE calculations, aside from possible issues related to differences in line shapes.

The original 3D hydrodynamic simulations were computed on a staggered Cartesian grid with  $240^3$  elements, extending deep below the stellar photosphere. However, for the radiative post-processing, we interpolate the hydrodynamic simulations to  $80 \times 80$  volume elements in the horizontal direction and 220 volume elements along the vertical and truncate the deep layers. This interpolation retains the refined resolution of the continuum-forming regions that have steep temperature gradients, as well as the optically thin line-forming regions. The truncation and interpolation follow that described in section 2.1 of Amarsi et al. (2018). The total number of volume elements is therefore of the order of  $10^6$ .

### 2.2.2 1D hydrostatic stellar model atmospheres

For comparison, we also use a set of custom 1D hydrostatic stellar model atmospheres which treat convection using the classical mixing length theory computed with the ATMO code written by W. Hayek (see appendix A of Magic et al. 2013 for details). The 1D models were computed at exactly the same stellar parameters as the 3D hydrodynamic models, share the same equation of state and opacities, and were based on the same opacity binning procedure. We also compute additional 1D models at randomly selected values of  $T_{\text{eff}}$  and  $\log g$  in order to test our interpolation procedures.

Unlike the 3D hydrodynamic models, where convective motions arise naturally from first principles, convection in these 1D models is approximated using mixing length theory controlled by a mixing length parameter ( $\alpha_{\text{MLT}} \equiv l/H_{\text{P}}$ , where  $l$  is the mixing length and  $H_{\text{P}}$  is the pressure scale height). We used models computed with  $\alpha_{\text{MLT}} = 1.0, 1.5$ , and  $2.0$ , thereby covering a range typically used in the literature. In the radiative transfer post-processing,

small-scale hydrodynamic motions that produce an effective broadening of line opacities leading to desaturation of spectral lines are mimicked through a local broadening parameter known as microturbulence (Asplund et al. 2000). We select the values  $v_{\text{mic}} = 0, 1, \text{ and } 2 \text{ km s}^{-1}$ . For our default comparisons, we use models with  $\alpha_{\text{MLT}} = 1.5$  and  $v_{\text{mic}} = 1 \text{ km s}^{-1}$ .

In addition to these custom model atmospheres, we also use the commonly used MARCS grid (Gustafsson et al. 2008), which covers a wide range in stellar parameters representative of FGKM-type dwarfs and giants.

### 3 RESULTS

In this section, we introduce the main NLTE effects on Li as discussed in the literature based on 1D (Section 3.1) and 3D (Section 3.2) modelling. We present our 3D line formation results in Section 3.3. In Section 3.4, we compare our results to previous work, and discuss a previously neglected effect of background line opacity blocking of lithium lines in the UV that would otherwise deplete low-lying populations through photon pumping. With this effect taken into account, we generally find stronger optical lines implying lower inferred lithium abundances. Lastly, in Section 3.5, we discuss the abundance correction over the HR diagram.

#### 3.1 Departures from LTE in 1D models

The main NLTE mechanisms relevant to lithium have been presented in a large body of previous studies (e.g. Steenbock & Holweger 1984; Carlsson et al. 1994; Pavlenko & Magazzu 1996; Takeda & Kawanomoto 2005; Lind et al. 2009; Osorio et al. 2011). While our model represents a substantial upgrade in terms of, e.g. collisional transition rates and the treatment of background opacities, the pedagogical description of NLTE mechanisms and behaviours by Carlsson et al. (1994) largely still applies to our results. As discussed by Lind et al. (2009), this agreement results from newer model atmospheres exhibiting steeper temperature gradients that increase the excess of ultraviolet radiation and thus leads to enhanced overionization, which partially cancels with the use of hydrogen collisional transition rates from Barklem et al. (2003) that are significantly larger than the previously used hydrogen collisional rates predicted through the Drawin (1968) formula. Furthermore, while Osorio et al. (2011) found large differences between their R-matrix calculations of electron collisional transition rates compared to previous semi-empirical estimates, the two happen to show agreement within a factor of 2 for the dominant 2s–2p resonance transition, resulting in negligible effects on level populations and abundance corrections.

When the 670.8-nm resonance line is strong, departures from LTE are mainly controlled by the line itself through photon losses in so-called resonance scattering. In NLTE, these photon losses dictate that the radiation field ( $J_\nu$ ) in the line becomes sub-thermal, i.e. smaller than the Planck function ( $B_\nu$ ). For a sufficiently strong line where line opacity dominates over continuum opacity, the source function ( $S_\nu$ ) becomes controlled by the line rather than the continuum. As the line source function ( $S_l$ ) becomes tied to the radiation field ( $S_l \approx J_\nu < B_\nu$ ), the drop in  $S_l$  causes core darkening that strengthens the line, thus leading to lower inferred abundances in NLTE compared to LTE.

The steep temperature dependence of  $B_\nu$  in the ultraviolet may produce a superthermal radiation field ( $J_\nu > B_\nu$ ), which has the opposite effect to resonance scattering. Photoionization through the continua of the ground and first excited states (2s and 2p) causes a drop in the populations of Li I through overionization in the presence

of steep temperature gradients or small ultraviolet opacities. The corresponding effect of photon pumping through ultraviolet spectral lines rather than continua is largely quenched by background metal-line opacities.

At infrared wavelengths, the opposite case will typically be true, with a sub-thermal radiation field ( $J_\nu < B_\nu$ ). The resulting deficit of photoionization of highly excited states leads to increased level populations through overrecombination. The flow of excess electrons propagates down through photon losses in a chain of infrared lines, and further to the ground state through the resonance line.

The interplay of these various NLTE mechanisms leads to abundance corrections that vary in magnitude and sign with stellar parameters as well as the abundance of lithium. A stronger resonance line will produce more negative abundance corrections through resonance scattering. Higher  $T_{\text{eff}}$ , and to lesser extent lower  $\log g$ , weakens the resonance line and further makes abundance corrections more positive through enhanced overionization. At lower [Fe/H] the smaller background opacities compete with a flattening of the atmospheric temperature gradients due to a lack of radiative cooling (in 1D model atmospheres), that in practice leads to only a small enhancement of overionization and thus abundance corrections that become slightly more positive.

The atom presented in Section 2.1 was also used to compute a comprehensive grid of 1D departure coefficients for use in GALAH DR3 (Buder et al., in preparation), and made available through Amarsi et al. (2020), and Amarsi (2020).

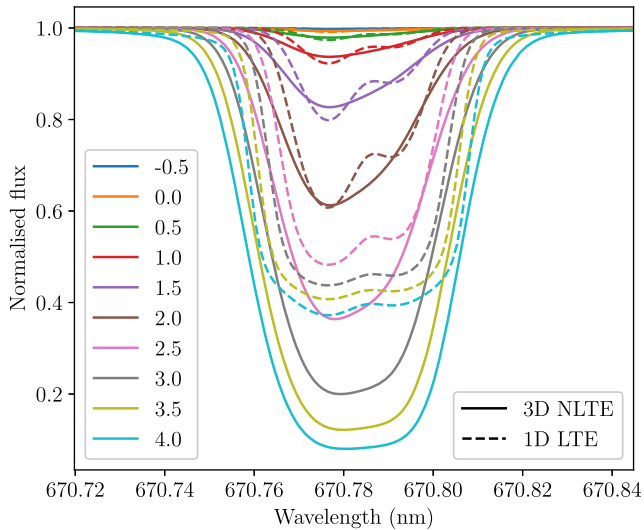
#### 3.2 Departures from LTE in 3D models

While the mechanisms outlined above are still the dominant ones in 3D model atmospheres, their relative importance and behaviour with stellar parameters changes due to the generally steeper temperature gradients present.

In particular, the outer layers of hydrodynamic atmospheres are effectively cooled by convective motions and adiabatic expansion. In warm metal-poor models, the deficit of opacities leads to radiative heating and cooling rates far below equilibrium, which alongside the convective cooling mechanism, results in significantly lower temperatures than seen in hydrostatic atmospheres in radiative equilibrium (Asplund et al. 1999; Collet, Asplund & Trampedach 2006). In LTE, low temperatures in the outer regions of the atmospheres lead to a decrease in the ionization fraction of lithium and thus strongly enhanced populations and resulting line strengths, in particular, in the steep temperature structures formed by hot upwelling gas in granules, producing negative abundance corrections of the order of 0.3 dex in warm metal-poor dwarfs (Asplund et al. 1999; Klevas et al. 2016).

Pioneering work utilizing limited radiative transfer in 3D hydrodynamic simulations of the Sun indicated that a decoupling of the radiation field from the local gas properties would drive significantly stronger overionization effects than in the corresponding 1D hydrostatic models, weakening the line by 0.1 dex (Kiselman 1997; Uitenbroek 1998). This decoupling was also observationally verified in resolved spectra of the solar surface (Kiselman 1998), which in contrast to 3D LTE predictions exhibited no significant variation of line strengths as a function of the continuum intensity. Calculations by Asplund et al. (2003) and Barklem et al. (2003) with the MULTI3D code (Botnen & Carlsson 1999) verified these early results, and extended work to metal-poor dwarfs where effects were significantly stronger with abundance corrections comparing 3D NLTE to 3D LTE of the order of 0.2 dex, resulting in good agreement between 3D NLTE and 1D NLTE calculations. They





**Figure 3.** Line profiles computed in 3D NLTE (solid lines) and 1D LTE (dashed lines) for a cool red giant star model with nominal parameters  $T_{\text{eff}} = 5000$  K,  $\log g = 2.0$ ,  $[\text{Fe}/\text{H}] = -2.0$ , and a range of  $A(\text{Li})$  (colour-coded as indicated in the legend). The 1D model uses  $\alpha_{\text{MLT}} = 1.5$  and  $v_{\text{mic}} = 1$  km s $^{-1}$ . Neither rotational nor macroturbulent broadening has been applied.

found that the effect of enhanced overionization was particularly pronounced in regions where the temperature gradient is strong, i.e. where LTE line strengths were the most strongly enhanced, while the opposite behaviour may appear above intergranular lanes.

Further work has extended the coverage of 3D NLTE calculations to better sample the spatial and temporal variations in the 3D hydrodynamic simulations of metal-poor dwarfs (Sbordone et al. 2010; Lind et al. 2013) and giants (Klevas et al. 2016; Nordlander et al. 2017) as well as solar-metallicity stars (e.g. Mott et al. 2017; Harutyunyan et al. 2018). We note, however, that these calculations are heterogeneous and the results are sometimes incompatible due to different implementations of model atoms (in terms of electron structure complexity and collisional rates), radiative transfer and opacity treatment, and model atmospheres.

### 3.3 3D NLTE line formation

We show in Fig. 3 profiles of the 670.8-nm resonance line for a very metal-poor model representing a star on the red giant branch. While the 1D LTE line profiles are sharp with clear fine-structure components, velocity fields in the 3D model of the order of 5 km s $^{-1}$  (as compared to the thermal linewidth of 3 km s $^{-1}$ ) cause a smoothed appearance for the 3D NLTE profiles. At the lowest  $A(\text{Li})$ , the line forms in relatively deep atmospheric layers where high pressures lead to high collisional rates that nearly uphold LTE. As  $A(\text{Li})$  increases, line formation begins to push higher in the photosphere toward layers with significantly lower temperatures in the hydrodynamic simulation than in the corresponding hydrostatic model. More importantly, however, photon losses in the core of the 670.8-nm line become significant, leading to a deeper line with an NLTE effect that strengthens with increasing  $A(\text{Li})$  through so-called resonance scattering.

Fig. 4 shows top-down views of four representative model atmospheres, as seen in disc-centre intensity. From the left- to right-hand side, these represent a star on the lower main-sequence, a metal-poor red giant branch star, the Sun, and a metal-poor main-sequence turn-off star as commonly observed on the Spite plateau. Each

row illustrates different properties of the 670.8-nm line: the local continuum intensity ( $I_{\text{cont}}$ ), the variation of log reduced equivalent width (REW) relative to its average in LTE ( $d\text{REW}_{\text{LTE}}$ ) and NLTE ( $d\text{REW}_{\text{NLTE}}$ ), and finally the NLTE effect on line strengths. In this study, we use the REW

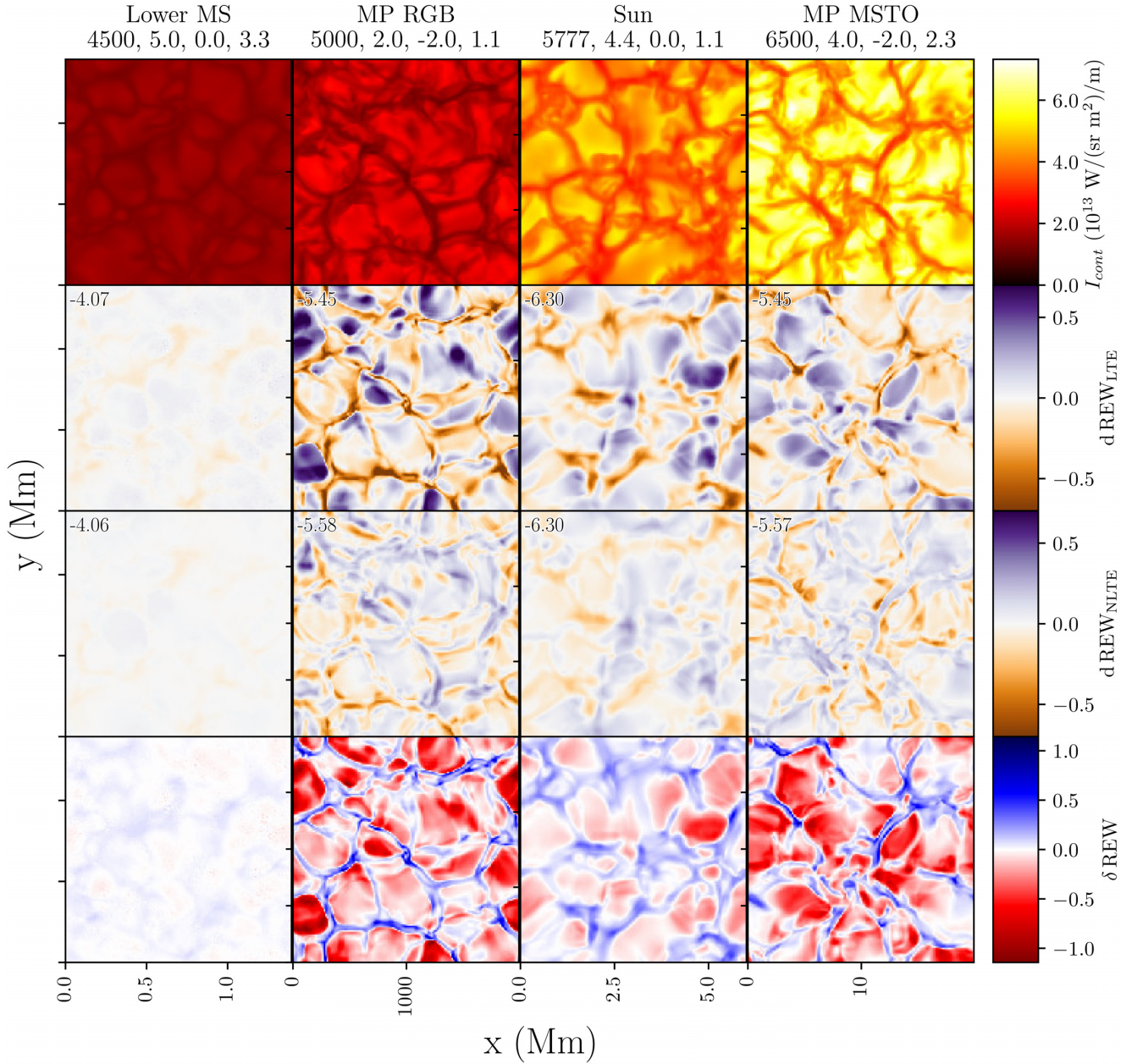
$$\text{REW} = \log_{10} \left( \frac{\text{EW}}{\lambda_0} \right), \quad (1)$$

where EW is the equivalent width of the line profile and  $\lambda_0$  its central wavelength at rest. The last row of Fig. 4 shows the difference in line strength between NLTE and LTE, defined as

$$\delta\text{REW} = \text{REW}_{\text{NLTE}} - \text{REW}_{\text{LTE}}. \quad (2)$$

While the dominant difference in continuum intensity is due to the variation in surface temperature and therefore overall brightness, differences, in contrast, can also be appreciated to increase with  $T_{\text{eff}}$  and decrease with  $\log g$  and  $[\text{Fe}/\text{H}]$  (for an in-depth discussion, see Magic et al. 2013). Like previous work, we find that the LTE line strength anticorrelates with the surface intensity pattern, such that bright upflowing granules due to their strong temperature gradients typically exhibit strong lines, compared to the cool intergranular lanes with weaker temperature gradients. In NLTE, however, this correlation is weakened due to the influence of non-local radiation fields. Foremost, the bright granules illuminate the line-forming layers immediately above, leading to a strongly superthermal radiation field. This drives an excess of photoionization in the continuum of the first excited state, 2p, leading to overionization that weakens the line strength relative to LTE by as much as a factor of ten for the two metal-poor models. In contrast, the line forms deeper above intergranular lanes, where the flatter temperature gradient does not produce a strongly superthermal radiation field, and instead, sub-thermal radiation across the infrared continua of more highly excited states drive overrecombination. A similar but milder effect of this is also seen in the solar model, where its weak line strength overall shifts line formation deeper. At the same time, its high-surface gravity and strong metal-line opacity both lessen all departures from LTE.

The discussion above focused mainly on the three warm models. The cool lower main-sequence model exhibits similar effects to the others, albeit at smaller magnitude due to the immense line strength of nearly 600 mÅ, because its equivalent width is proportional to  $\sqrt{n}$  ( $n$  is the population of lower state) instead of proportional to  $n$  like for weak lines. The combination of low surface temperature and high  $A(\text{Li})$ , however, means the subordinate lines at 610.4 and 812.6 nm become measurable. We therefore look more closely at this model by examining its surface variation for all three lithium lines in Fig. 5. The subordinate lines originate from the first excited state at  $E_{\text{low}} = 1.84$  eV and are therefore significantly weaker than the resonance line, and thus probe different heights of formation as compared to the resonance line. For the 670.8-nm line, the core depth is mainly controlled by resonance scattering, which compared to LTE always enhances the line strength through core darkening. This results in the line core having similar depth across the entire surface in NLTE. In LTE, however, the core may be significantly weakened either due to the weak temperature gradient seen above intergranular lanes, or due to temperature contrast inversions in the optically thin layers caused by mechanical heating. In NLTE, the wings of the 670.8-nm line as well as the cores of the 610.4- and 812.6-nm lines, are instead controlled by the excess of UV flux. There is a balance between overionization and overrecombination that is driven in a similar way to that seen for the other models in Fig. 4, and where on average overionization dominates.



**Figure 4.** Top-down view of different 3D stellar models (columns), illustrating different properties of the 670.8-nm transition in disc-centre intensity (rows). Models are labelled by their stellar parameters,  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$  and  $A(\text{Li})$ . From the top to bottom panels, the rows illustrate the continuum intensity across the stellar surface at 670.8 nm ( $I_{\text{cont}}$ ), the variation in LTE REW ( $\text{REW}_{\text{LTE}}$ ) relative to its average ( $d\text{REW}_{\text{LTE}}$ ), likewise in NLTE ( $d\text{REW}_{\text{NLTE}}$ ), and the difference between  $\text{REW}_{\text{NLTE}}$  and  $\text{REW}_{\text{LTE}}$  ( $\delta\text{REW}$ ). The average values of  $\text{REW}_{\text{LTE}}$  and  $\text{REW}_{\text{NLTE}}$  are labelled in the top left of the corresponding panels.

### 3.4 The role of UV lines

In order to quantify NLTE effects, we define abundance corrections,

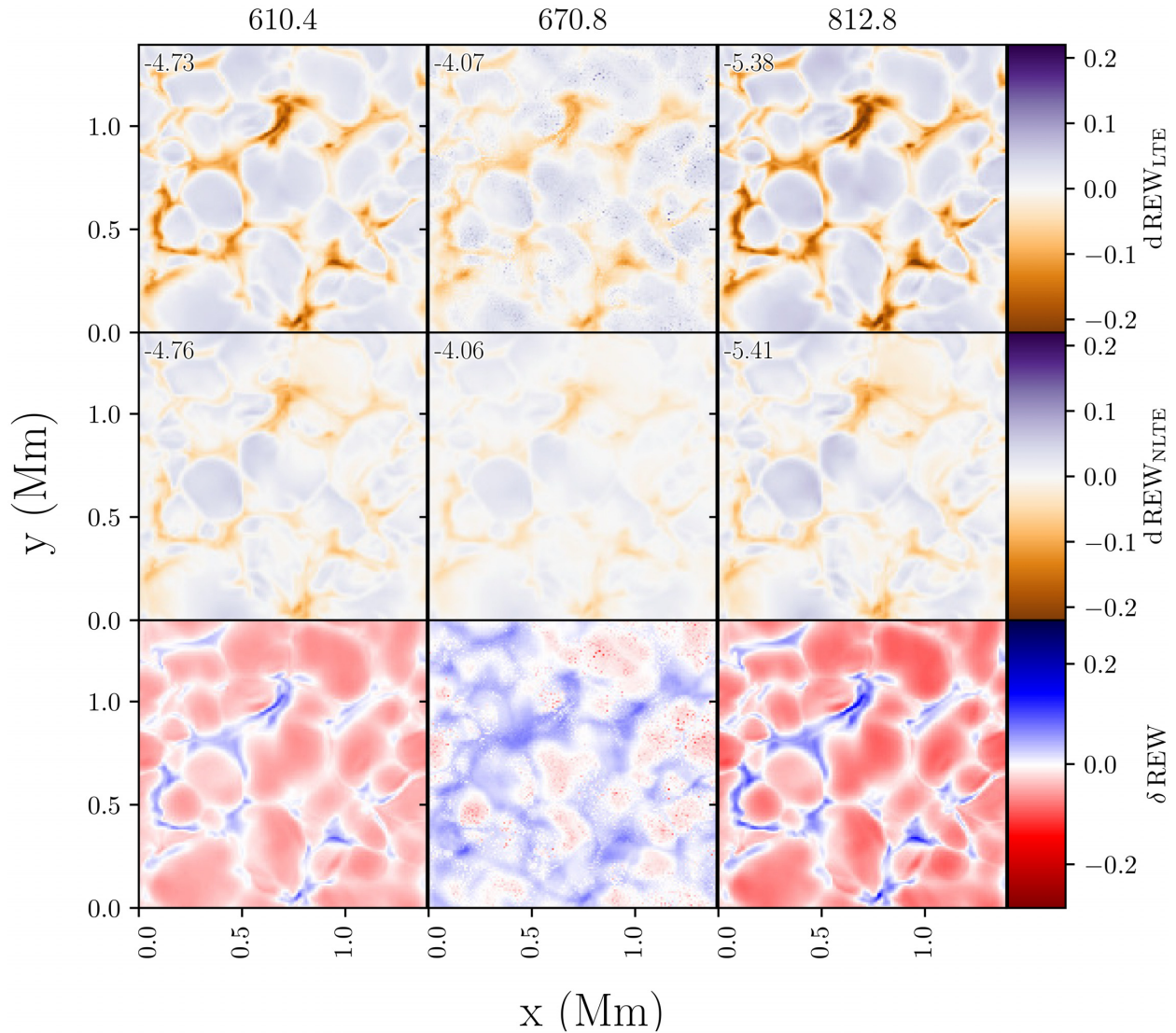
$$\Delta\text{NLTE} = A(\text{Li})_{3\text{D,NLTE}} - A(\text{Li})_{1\text{D,LTE}}, \quad (3)$$

where  $A(\text{Li})$  is calculated based on matching REW. Our  $\Delta\text{NLTE}$  are generally more negative than found in previous work, as shown in Fig. 6, with differences being larger at lower  $T_{\text{eff}}$  and  $\log g$ , and higher  $[\text{Fe}/\text{H}]$ . This is an effect of our more accurate treatment of background line opacities: In BALDER, overlapping transitions are implemented through the use of a common set of wavelengths and opacities for all line and continuum transitions, including the background lines and continua from other elements that are treated

in LTE. After computing the statistical equilibrium, background line opacities can optionally be switched on or off in order to produce emergent spectra with or without blends, while overlapping transitions in the atom of study are always retained unless individual transitions are explicitly disabled. This is in contrast to, e.g. the MULTI code (Carlsson 1986), which does not treat overlapping transitions and by default only considers background line opacities that blend with continua but not line transitions (see e.g. Collet, Asplund & Thévenin 2005), unless blending line transitions are specified explicitly (e.g. as was done by Nordlander & Lind 2017).

We illustrate effects of partially disabling background line opacities in Fig. 7, where we compare results from Lind et al. (2009) to test calculations with MARCS models using the same model atom as





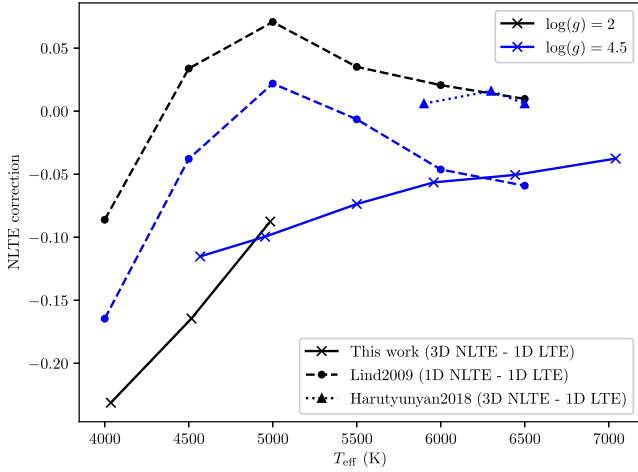
**Figure 5.** Top-down view of a 3D stellar model with  $T_{\text{eff}} = 4500\text{ K}$ ,  $\log g = 5.0$  and  $[\text{Fe}/\text{H}] = 0.0$ , calculated with  $A(\text{Li}) = 3.3$  for different spectral lines (columns, labelled by wavelength in units of nm). From the top to bottom panels, the rows illustrate the variation in LTE REW ( $\text{REW}_{\text{LTE}}$ ) relative to its average ( $d\text{REW}_{\text{LTE}}$ ), likewise in NLTE ( $d\text{REW}_{\text{NLTE}}$ ), and the difference between  $\text{REW}_{\text{NLTE}}$  and  $\text{REW}_{\text{LTE}}$  ( $\delta\text{REW}$ ). The average values of  $\text{REW}_{\text{LTE}}$  and  $\text{REW}_{\text{NLTE}}$  are labelled in the top left-hand side of the corresponding panels.

in their work. As expected, we find that removing background line blocking for lines redward of 600 nm mainly causes a decrease of photon losses in the 670.8-nm line, which is itself normally responsible for increasing the ground-state population through so-called resonance scattering. When disabling background line blocking of all line transitions, we find that a number of UV resonance and subordinate lines together deplete the 2s and 2p states through photon pumping by an amount comparable to that of UV photoionization, resulting in a significant weakening of all low-excitation lines. In contrast, through our consistent inclusion of background line opacities, we find that radiative excitations through UV line transitions are effectively quenched, having radiative brackets (the net difference between in-going and out-going radiative transition rates) that are effectively zero. We note that Carlsson et al. (1994) found that photon pumping through UV lines was not important – most likely this is due to their use of older model atmospheres with generally flatter temperature structures that produced less superthermal radiation.

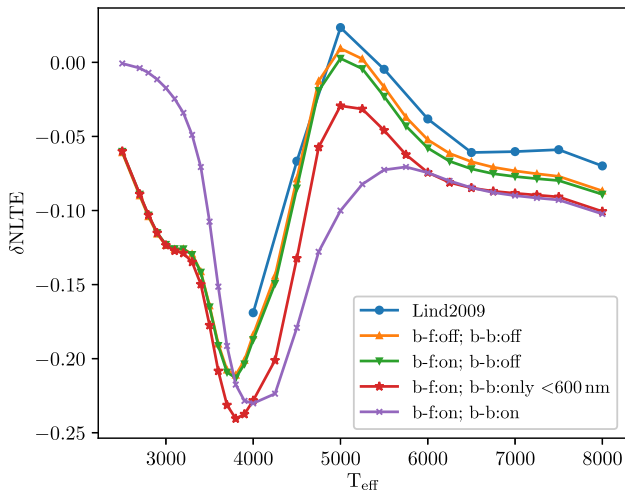
### 3.5 3D NLTE abundance corrections

Fig. 8 shows  $\Delta\text{NLTE}$  for the 670.8-nm line at a representative abundance  $A(\text{Li}) = 2$ , for four select values of  $[\text{Fe}/\text{H}]$ . In each panel, we show results for our 3D hydrodynamic models, as well as an interpolation based on a method that is explained further in Section 4.2, to highlight how departures from LTE vary with stellar parameters.

Overall, we find that  $\Delta\text{NLTE}$  tends to become more positive with increasing  $T_{\text{eff}}$  and decreasing  $[\text{Fe}/\text{H}]$ , but depends only weakly on  $\log g$ . At low  $T_{\text{eff}}$ , the line is sufficiently strong that resonance line scattering and overrecombination are both important NLTE effects, together these effects cause the line to increase in strength and therefore yield lower  $A(\text{Li})$ . As  $T_{\text{eff}}$  increases, the line strength and therefore the effect of resonance line scattering decreases while at the same time overionization becomes more important. Both effects lead to a weaker line, and thus more positive  $\Delta\text{NLTE}$ . With decreasing line strength, line formation moves inward toward



**Figure 6.** NLTE correction for models with varying  $T_{\text{eff}}$ , comparing this work (3D NLTE–1D LTE, solid line with crosses), Lind et al. (2009) (1D NLTE–1D LTE, dashed line with circles), and Harutyunyan et al. (2018) (3D NLTE–1D LTE, dotted lines with triangles). All models use  $[\text{Fe}/\text{H}] = -1$  and  $A(\text{Li}) = 2$ . We show models with  $\log g = 2.0$  (in black) and  $\log g = 4.5$  (in blue).



**Figure 7.** Test calculations illustrating the 1D NLTE abundance correction ( $\delta\text{NLTE}$ ) for MARCS models with  $\log g = 4.0$ ,  $[\text{Fe}/\text{H}] = -1.0$ , and  $A(\text{Li}) = 2.0$ . Models have been computed using the model atom from Lind et al. (2009) with background line opacities enabled or disabled for bound-free (b-f) and line (b-b) transitions of lithium (colours according to figure legend); results from Lind et al. (2009) are shown for comparison (in blue).

deeper layers, where overionization becomes less important relative to mutual neutralization (i.e. recombination) through charge transfer, and the trend therefore turns over when  $T_{\text{eff}} \sim 6000$  K in the most metal-poor models. At higher  $[\text{Fe}/\text{H}]$ , increasing opacities tend to quench overionization while at the same time higher gas and electron pressure produce higher collisional rates, both of which lead to smaller departures from LTE. Variations with  $\log g$  are very small due to a cancellation effect: the higher pressure leads to increased collisional rates and thus more efficient thermalization, as well as a decreasing ionization fraction which strengthens the resonance line and may thus increase the effect of resonance scattering.

We stress that while abundance effects driven by equivalent width differences are a helpful tool to understand departures from LTE and indeed to quickly apply corrections to existing 1D LTE abundance

analyses, the line shapes themselves also change. As shown in Fig. 3, while weaker lines often exhibit similar strength in 3D NLTE and 1D LTE leading to small  $\Delta\text{NLTE}$ , saturated and strong lines may differ dramatically in both equivalent width and shape. Using  $\Delta\text{NLTE}$  to derive 3D NLTE  $A(\text{Li})$  may be further hampered by the dependence of 1D LTE analyses on fudge factors such as  $\alpha_{\text{MLT}}$  and  $v_{\text{mic}}$ .

#### 4 INTERPOLATION

In this section, we show that straightforward spline interpolation of our grid of synthetic spectra yields a larger average error in  $A(\text{Li})$  compared to other interpolation methods. We therefore investigate more involved interpolation methods, in order to provide to the community a package for accurate abundance analyses based on our synthetic 3D NLTE spectra computed with the BALDER code: BREIDABLIK.<sup>3</sup>

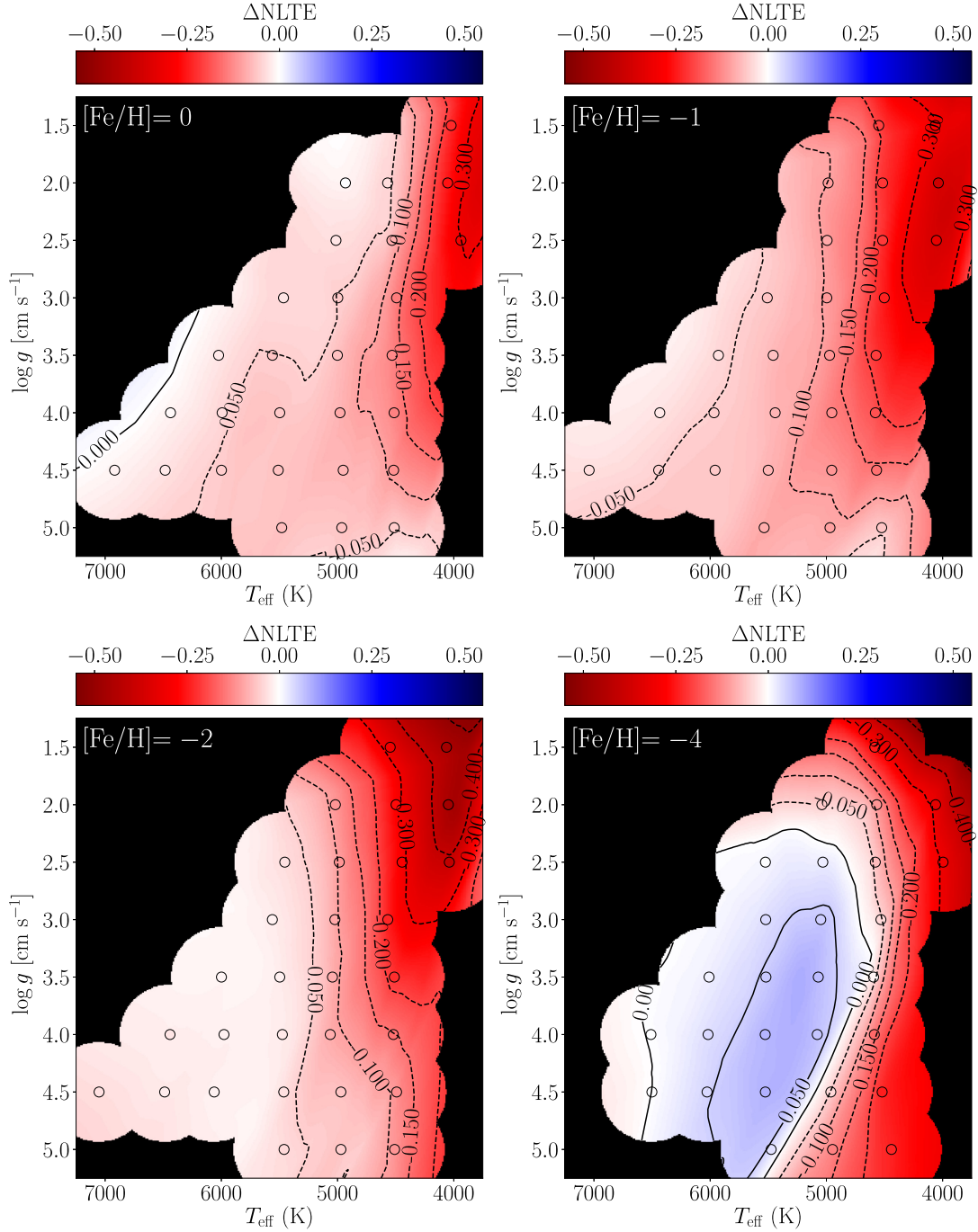
Interpolation for this grid is complicated due to the fact that  $T_{\text{eff}}$  is an output rather than input parameter in 3D hydrodynamic simulations. Therefore, the grid is tabulated at irregular intervals in this dimension, with values that are slightly different for every  $\log g$  and  $[\text{Fe}/\text{H}]$ . As a result, interpolation methods requiring data points at regular intervals cannot be used out of the box. Additionally, our grid of model atmospheres has relatively large steps in parameter space (see Fig. 2), resulting in naturally higher errors in interpolation.

There exist many non-linear interpolation methods which can be used to predict abundances from an observed spectrum using an irregular grid. Many of these use inverse modelling, where stellar parameters are predicted from observed spectra rather than the other way around. For example, Snider et al. (2001) used artificial neural networks with back-propagation, and applied this to medium-resolution spectra to predict stellar parameters; The Gaia Data Processing and Analysis Consortium uses extremely randomized trees (Geurts, Ernst & Wehenkel 2006), which is an ensemble method that predicts stellar parameters on the parameter space spanned by the training set (Andrae et al. 2018); StarNet uses a convolutional neural network applied to APOGEE spectra to predict stellar parameters and 15 elemental abundances in an inverse model (Fabbro et al. 2017); AstroNN similarly uses a Bayesian neural network with dropout variational inference applied to APOGEE spectra to predict stellar parameters and 18 elemental abundances, with error estimates (Leung & Bovy 2019). The Cannon (Ness et al. 2015; Casey et al. 2016; Ho et al. 2016) uses a polynomial model to determine stellar parameters and elemental abundances from a spectrum. However, unlike the other models, the Cannon can also be applied to forward modelling, where it generates a spectrum given stellar parameters and elemental abundances. Another forward modelling approach is the Payne (Ting et al. 2019), which uses a fully connected feedforward neural network to rapidly and accurately predict spectra given stellar parameters and abundances, followed by a fit to APOGEE spectra to determine stellar parameters and 20 elemental abundances.

While the inverse models have certain advantages, e.g. in terms of speed and always returning an answer, forward modelling is more flexible. For BREIDABLIK, we therefore choose a forward modelling approach, and provide tools that fit our synthetic spectra to observations. This has the advantage that our method is not tied to any particular instrumental setup, but can be implemented as part of any traditional analysis. It can also be further integrated into more advanced methodologies, that evaluate systematic errors or advanced statistics including deriving upper limits to non-detections.

<sup>3</sup><https://github.com/ellawang44/Breidablik>





**Figure 8.** 3D NLTE–1D LTE abundance corrections ( $\Delta\text{NLTE}$ ) for the 670.8-nm line, shown at a 1D LTE reference abundance of  $A(\text{Li}) = 2$ . Each panel shows a different metallicity, from left to right, top to bottom, these are:  $[\text{Fe}/\text{H}] = 0, -1, -2,$  and  $-4$ . The circles show the calculated  $\Delta\text{NLTE}$  from our models, whilst the surface shows the predicted  $\Delta\text{NLTE}$  which has been interpolated according to Section 4.2.

We investigate the interpolation of both spectral line profiles and line strengths. The interpolation of spectral line profiles is subtly different from that of line strengths as they involve the extra dimension of wavelengths. While line strengths vary smoothly across stellar parameter space, their *shape* is more complicated due to broadening processes and hydrodynamic velocity fields that may redistribute flux between adjacent wavelength points in a non-linear way.

We test a number of different interpolation methods for both line profiles and strengths, and evaluate them using leave-one-out cross-validation. Hyperparameters for each method are optimized through five-fold cross-validation, as per standard practice (see chapter 5 of James et al. 2013; chapter 4 of Kuhn & Johnson 2013), using a common seed to generate the folds. As spectrum variations across stellar parameters are more difficult to predict than variations with  $A(\text{Li})$ , folds were created based on  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  only. The

final model is trained using the regular models in the STAGGER grid, and validated on models tailored to particular stars, that were not part of the training set. Sections 4.1 and 4.2 discuss interpolating spectral line profiles and line strengths, respectively, presenting and comparing the different interpolation methods tested.

#### 4.1 Interpolating spectral line profiles

We test and compare three interpolation methods: spline interpolation, the Cannon, and Kriging. Our implementation of the different methods and the hyperparameters used are described in Sections 4.1.1–4.1.3.

In order to make the normalized flux easier to interpolate, especially for very weak lines, we transform it to a quantity that scales better with abundance. On the weak part of the curve of growth, the flux depression varies approximately linearly with the number of absorbers, so we define a transformed flux

$$f_t = \log_{10}(1 - f + s), \quad (4)$$

where  $f$  is the normalized flux, and  $s$  is a small positive constant (treated as a hyperparameter) which smoothly truncates  $f_t$  as the flux approaches the continuum. Fig. 9 illustrates how the normalized flux compares to the quantity  $f_t$ , for two representative values of  $s$ . Importantly,  $f_t$  varies nearly linearly with  $A(\text{Li})$  for small flux depressions, with a dynamic range controlled by  $s$ , before reaching saturation as the flux approaches zero. Conversely, as the flux approaches the continuum we find  $f_t = \log_{10}s$ , which helps avoid numerical noise dominating the model. Test runs where we interpolate over the normalized flux  $f$  tend to incur larger relative errors when the flux depression is small, corresponding to larger abundance errors and larger errors in the detailed shape for very weak lines.

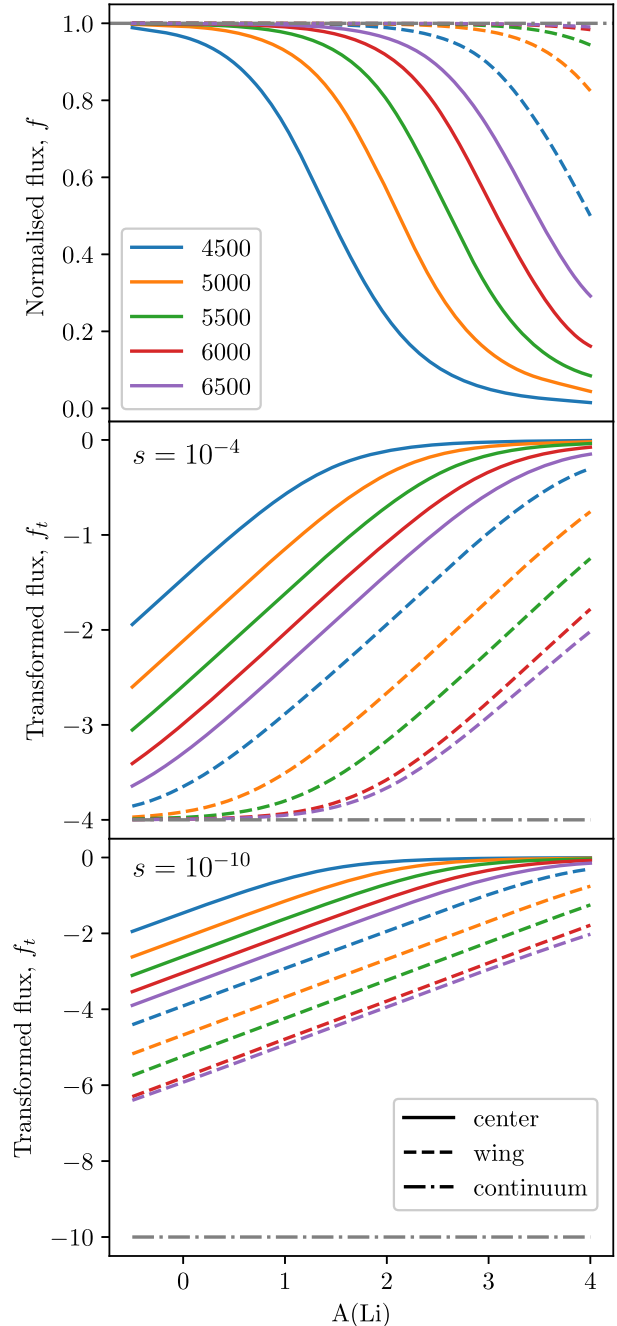
All interpolation methods are trained on all 3 Li lines and on a per pixel basis.

##### 4.1.1 Spline interpolation

Local polynomial models like spline interpolation are the most straightforward, and most commonly used in the literature. Due to the irregular spacing of  $T_{\text{eff}}$  in our grid of 3D models, we perform a “training” step following Amarsi et al. (2018), wherein we create a finer grid with regular steps in stellar parameters  $T_{\text{eff}}$ ,  $\log g$  and  $[\text{Fe}/\text{H}]$  using cubic spline interpolation and linear extrapolation. We then use trilinear interpolation within this finer grid to interpolate to the exact stellar parameters at each  $A(\text{Li})$ , and produce the final spectrum using cubic splines over  $A(\text{Li})$ . This method uses only a single hyperparameter, which we optimise to  $s = 10^{-9}$  (see equation 4).

##### 4.1.2 The Cannon

The Cannon is a global polynomial interpolation method (Ness et al. 2015; Ho et al. 2016). We provide the Cannon with three labels:  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ , training one model per  $A(\text{Li})$ , then using cubic spline over  $A(\text{Li})$  to compute the final interpolated line profile. The Cannon is normally trained by applying a  $\chi^2$  minimization method to observed data. Since our synthetic data are not stochastic, it therefore does not carry statistical uncertainties, and as such, we set the variance in pixel to a constant value. Our hyperparameter test finds best results when using a third-order polynomial form with 20 terms (including cross-terms) describing the flux in each pixel, and  $s = 10^{-9}$ .



**Figure 9.** Comparison of the normalized flux and how it varies as a function of  $A(\text{Li})$  (top panel), to the transformed flux  $f_t$  (middle and bottom panels) for two different values of the softening parameter  $s$ . Solid lines show the flux at line centre, dashed lines represent a point in the line wing (offset by  $-0.03$  nm from the core), and the dash-dotted line is the continuum. Models of different  $T_{\text{eff}}$  are shown with different colour; all models use  $\log g = 4$  and  $[\text{Fe}/\text{H}] = -2$ .

##### 4.1.3 Kriging

Kriging interpolates by computing the distance-dependent weighted averages of points in a neighbourhood. We define this neighbourhood as the entire grid to remove discontinuities. We use the `OrdinaryKriging3D` class from the Python package `PyKriging` to interpolate across stellar parameters individually at each  $A(\text{Li})$ , and

**Table 1.** Comparison of different interpolation methods used to predict the 670.8-nm line profile, showing their MAD and RMS error statistics in predicting abundance for leave-one-out cross-validation on the 3D grid, and interpolation and extrapolation tests on a 1D grid. Also shown is the time taken to train the models for all three Li lines (418 pixels total), and the time it takes to execute the model to produce interpolated line profiles for all three lines with arbitrary values of  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $\text{A}(\text{Li})$ .

Method	3D leave one out		1D interpolation		1D extrapolation		Training time (s)	Execution time (s)
	MAD	RMS	MAD	RMS	MAD	RMS		
Spline	0.046	0.115	0.021	0.060	0.028	0.219	1880	0.0941
The Cannon	0.020	0.029	0.085	0.176	0.062	0.088	11.1	0.138
Kriging	0.014	0.022	0.020	0.038	0.019	0.143	235	5.62

then interpolate over  $\text{A}(\text{Li})$  by cubic splines to produce the final spectrum. Our best hyperparameter is  $s = 10^{-7}$ .

#### 4.1.4 Errors: 3D leave one out

We summarize the abundance-fitting statistics over the 670.8-nm line for the three investigated models in Table 1. Our error statistics are based on the median absolute deviation (MAD) and root mean square error (RMS) measures.

For spline interpolation, we find that models located near the edges of the grid incur very large errors that drive a large RMS error statistic. This is the expected behaviour for spline models, as they tend to extrapolate poorly. A particular difficulty is the fact that our grid does not have regular edges, but the upper and lower limits on  $T_{\text{eff}}$  vary with  $\log g$ . In contrast, we find that both The Cannon and Kriging performs very well on average, and are relatively reliable also when extrapolating to edge models. We select the interpolation method with the smallest MAD error statistic, Kriging, to use in BREIDABLIK. We opt for the MAD statistic rather than the RMS error as the former is less sensitive to outliers, and more representative of the typical expected error.

Table 1 shows also representative values for the training time and execution time for each interpolation method, as executed on a single CPU core on a modern desktop computer. The training time represents the execution time required to determine the model with our optimum set of hyperparameters, and the execution time is the time required to predict the lithium abundance for a single spectral line. The Cannon is by far the fastest in terms of training time due to its use of weighted least-squares fitting and its simple polynomial form. In comparison, the training time of Kriging and spline interpolation methods are slower by several orders of magnitude.

#### 4.1.5 Errors: verification models

To verify the validity of the final interpolation models, and to show the errors expected in practice, we used a verification set of four models that are not part of the regular STAGGER grid but rather were tailored to individual stars. The exact line profiles computed with BALDER for these verification models are compared to the interpolated line profiles from the three tested interpolation methods in Fig. 10, and the measured  $\text{A}(\text{Li})$  using each interpolation method is reported in Table 2. The interpolated line profiles have a similar shape to the synthesized profiles and replicates the characteristic asymmetry of the Li line. Overall, we find that spline interpolation and Kriging both perform well. Whilst the Cannon has been used successfully in many different cases, mainly in applications to large sets of empirical spectra (Ness et al. 2015; Ho et al. 2017; Buder et al. 2018), we note that the Cannon does not perform as well as other interpolation methods for our particular data set due to the large span in stellar parameter space. Compared to Table 1, all verification model errors

are notably smaller than the measured leave-one-out errors. As all four verification models are located either near regular models or well inside the grid boundaries, spline, and Kriging perform much better on the verification models compared to the measured MAD error statistic in the leave-one-out tests. These verification results indicate that the final interpolation methods produce consistent results when compared to the leave-one-out cross-validation errors.

#### 4.1.6 Errors: 1D comprehensive grid

Leave-one-out cross-validation errors presented in Section 4.1.4 likely yields overestimated error estimates. This is because the effective interpolation step in these tests will be 500 K in  $T_{\text{eff}}$ , 0.5 dex in  $\log g$  and 0.5–1 dex in  $[\text{Fe}/\text{H}]$ , while in practice interpolation on the full grid will require steps of at most 250 K in  $T_{\text{eff}}$ , 0.25 dex in  $\log g$  and 0.25–0.5 dex in  $[\text{Fe}/\text{H}]$ . While the tests on verification models presented in Section 4.1.5 yield fair estimates of the interpolation error, they cover only a very limited parameter space. Due to the large computational cost of creating 3D hydrodynamic simulations, it is beyond our computational resources to generate additional verification models that cover the entire parameter space for comparison to our 3D NLTE calculations. In this section, we present errors from a more comprehensive grid of verification models based on 1D NLTE calculations.

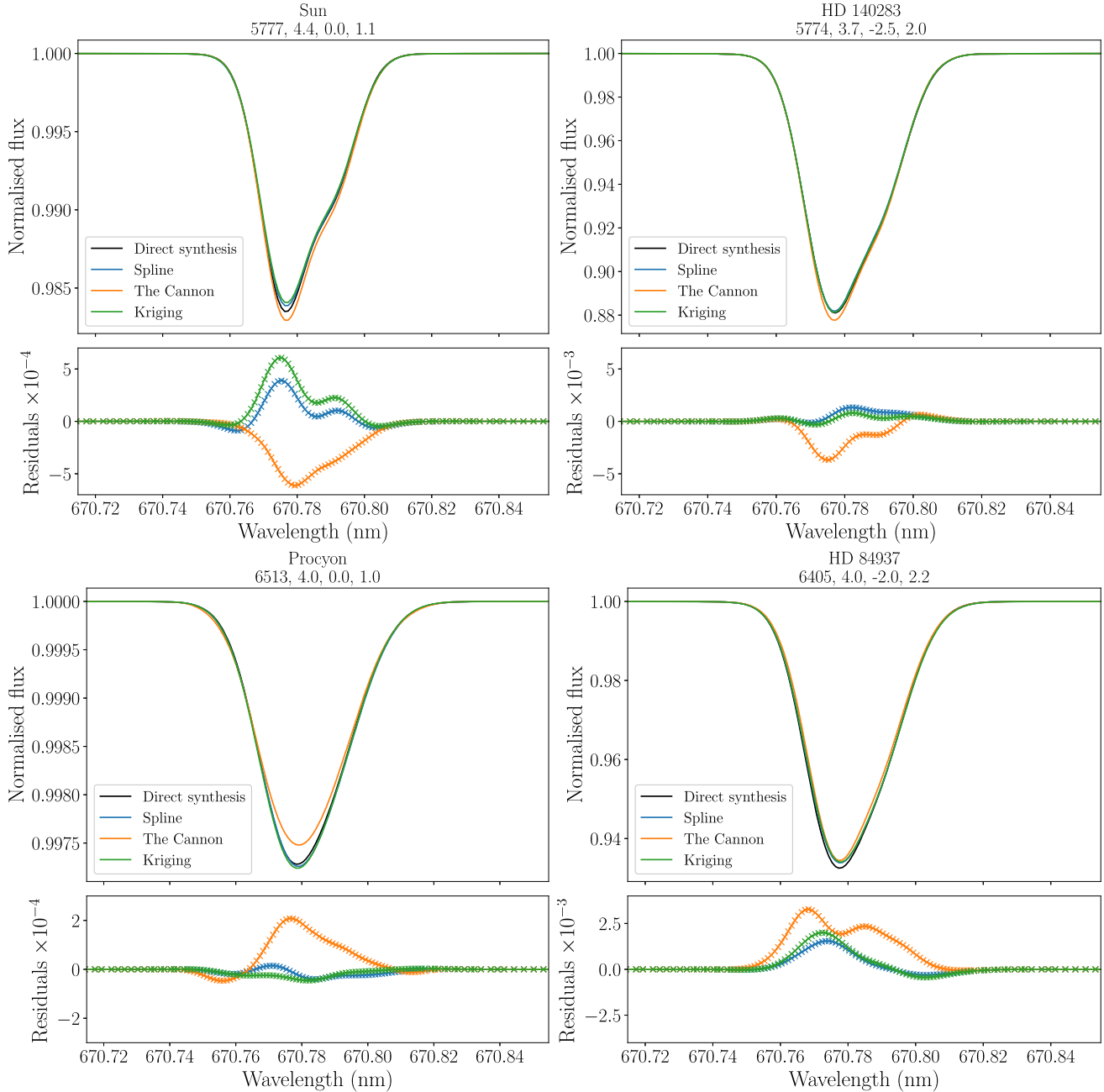
We use spectra computed using 1D hydrostatic models that were computed with the ATMO code, adopting  $\alpha_{\text{MLT}} = 1.5$ ,  $v_{\text{mic}} = 1 \text{ km s}^{-1}$ . We train spectrum interpolation models using the same hyperparameters as were used for our 3D NLTE spectra. We test on a total of 2000 1D ATMO models with randomly generated values for  $T_{\text{eff}}$  and  $\log g$ , while discrete values of  $[\text{Fe}/\text{H}]$  are selected from the existing grid, and  $\text{A}(\text{Li})$  is drawn from a normal distribution with mean 2.5 and standard deviation 0.3. In total, 1000 of these models are bounded by the existing grid models, hence are “interpolation” models; the other 1000 models are not fully bounded by existing grid models, hence are “extrapolation” models.

The MAD and RMS errors determined for these two samples are shown in Table 1. We find that spline interpolation performs well on average on both interpolation and extrapolation, while a small number of extreme outliers in the extrapolation sample produce a large RMS error. The Cannon surprisingly performs better on extrapolation than interpolation, indicating that there are likely small oscillations in the model. Our preferred model, Kriging, performs well on both interpolation and extrapolation, again with a small number of extreme outliers producing a large RMS error in the extrapolation sample.

## 4.2 Interpolating line strengths

As a complement to our interpolation of spectral line profiles, we also provide Python classes in BREIDABLIK to derive abundances





**Figure 10.** Comparison of interpolation models with verification models (that were not trained on). The verification models were tailored to specific stars, indicated by name and stellar parameters:  $T_{\text{eff}}$ ,  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $A(\text{Li})$ . Note that the residuals between prediction and verification model have been magnified by a different amount in each plot.

and abundance corrections based on REWs for all three Li lines. Our method for providing abundance corrections,  $\Delta\text{NLTE}$ , takes as input stellar parameters and  $A(\text{Li})$  measured in 1D LTE. Our method for providing 3D NLTE abundance measurements,  $A_{\text{REW}}$ , takes as input stellar parameters and a REW (equation 1).

We test a total of five interpolation models, including the spline interpolation, the Cannon and Kriging, presented in Sections 4.1.1–4.1.3, which are used here with no need for hyperparameters because the flux transform is not used. We introduce here two additional interpolation methods: multilayer perceptron (MLP) and support vector regression (SVR). These methods were also evaluated in

our tests for spectral line interpolation, but we found that due to limited precision and the random nature of the training procedure, both methods produce pixel-to-pixel variations that resemble noise in the line profiles. As a result, MLP and SVR are not considered for interpolating line profiles.

We note that Mott et al. (2020) present a method similar to our application of the Cannon, where they use a polynomial fit over  $T_{\text{eff}}$ ,  $\log g$ , and  $A(\text{Li})$ , and derive fitting functions individually at each value of  $[\text{Fe}/\text{H}]$ . They test their fitting functions on their input data, and find that errors are typically in the range of 0.020 dex for both  $\Delta\text{NLTE}$  and  $A_{\text{REW}}$ .

**Table 2.** Error in the predicted A(Li) using line profile interpolation models on verification models (that were not trained on). The input A(Li) used is shown in the first row, labelled ‘direct synthesis’, while subsequent rows show the error relative to the reference. The verification models were tailored to specific stars, indicated by name, whose stellar parameters are given in Fig. 10.

Model Method	Sun	HD 140283	Procyon	HD 84937
Direct synthesis	1.100	2.000	1.000	2.200
Spline	0.006	0.004	−0.004	0.007
The Cannon	−0.015	−0.010	0.026	0.019
Kriging	0.012	0.002	−0.006	0.009

**Table 3.** Comparison of different interpolation methods used to predict the abundance correction ( $\Delta$ NLTE) and abundance based on equivalent widths ( $A_{\text{REW}}$ ), for the 670.8-nm line. The columns are the same as in Table 1.

Line strengths	Interpolation	3D leave one out		1D interpolation		1D extrapolation		Training time (s)	Execution time (ms)
		MAD	RMS	MAD	RMS	MAD	RMS		
$\Delta$ NLTE	Spline	0.040	0.083	0.012	0.139	0.020	0.209	29.3	0.578
	The Cannon	0.025	0.041	0.024	0.148	0.027	0.199	0.198	0.002 63
	Kriging	0.021	0.037	0.012	0.142	0.012	0.170	0.567	15.1
	SVR	0.017	0.027	0.018	0.140	0.014	0.163	27.1	0.119
	MLP	0.012	0.020	0.012	0.140	0.011	0.173	2140	0.384
$A_{\text{REW}}$	Spline	0.051	0.663	0.026	0.063	0.024	0.202	28.3	0.563
	The Cannon	0.038	0.066	0.026	0.057	0.020	0.069	0.273	0.002 09
	Kriging	0.013	0.022	0.025	0.046	0.018	0.149	0.552	14.9
	SVR	0.018	0.036	0.030	0.056	0.034	0.080	1270	0.170
	MLP	0.010	0.014	0.027	0.047	0.015	0.073	133	0.176

#### 4.2.1 Support vector regression

SVRs is a form of support vector machines (Vapnik 1995), that defines a subset of the training set (the support vectors) which fall within  $\varepsilon$  of the geometric margin separating data, using a particular kernel function to transform the input space. We use the SVR class (Chang & Lin 2011) from the Python package *scikit-learn* (Pedregosa et al. 2011). For  $\Delta$ NLTE, we find best hyperparameters: the penalty parameter,  $C = 100$ ; penalty distance,  $\varepsilon = 10^{-3}$ ; and kernel function,  $f_{\text{ker}} =$  radial basis function (rbf). For  $A_{\text{REW}}$ , we find best hyperparameters:  $C = 5000$ ,  $\varepsilon = 10^{-5}$ , and  $f_{\text{ker}} =$  rbf.

#### 4.2.2 Multilayer perceptron

MLPs are fully connected feed forward neural networks, that connect “neurons” in a series of layers through a mixture of linear and non-linear transforms. We use the MLPRegressor class (Hinton 1990) from the Python package *scikit-learn* (Pedregosa et al. 2011). We set the maximum number of iterations to  $10^5$  and the tolerance to  $10^{-6}$ . For  $\Delta$ NLTE, our best hyperparameters are: number of layers,  $n_l = 2$ ; the number of neurons per layer,  $n = 900$ ; the L2 penalty,  $\alpha = 0.1$ ; and the activation function,  $f_{\text{act}} =$  rectified linear unit (ReLU). For  $A_{\text{REW}}$ , the best hyperparameters are  $n_l = 2$ ,  $n = 350$ ,  $\alpha = 0.01$ , and  $f_{\text{act}} =$  ReLU.

#### 4.2.3 Errors

We show the fitting statistics MAD and RMS errors of  $\Delta$ NLTE and  $A_{\text{REW}}$  models in Table 3. These reported errors are also likely to be upper bounds on the errors expected from fully trained grids, similar to Table 1. SVR takes much longer to train over  $A_{\text{REW}}$  compared to  $\Delta$ NLTE because the training time is sensitive to the hyperparameter,  $C$ , where larger values of  $C$  tends to takes longer to train due to

higher numbers of iterations required. MLP takes longer to train over  $\Delta$ NLTE compared to  $A_{\text{REW}}$ , because the training time is sensitive to  $n$ , as larger  $n$  implies a higher number of fitted parameters. Overall, our preferred method, MLP, produces by far the lowest errors and executes rapidly.

To verify the validity of the final line strength interpolation models, we use the same verification set as in Section 4.1. The exact  $\Delta$ NLTE and 3D NLTE A(Li) computed with BALDER for these verification models are compared to the interpolated line strengths from the five tested methods in Table 4. The verification results for the Cannon, Kriging, SVR, and MLP in Table 4 mostly matches the results in Table 3. Spline interpolation and Kriging both perform significantly better for the verification models again due to the location of the verification models being near regular models or well inside grid boundaries. These verification results indicate that the final interpolation models for both  $\Delta$ NLTE and  $A_{\text{REW}}$  are consistent with the leave-one-out cross-validation error statistics.

We also compute errors for the  $\Delta$ NLTE and  $A_{\text{REW}}$  models on a 1D comprehensive grid, similar to Section 4.1.6. Since these models are trained on 1D hydrostatic model atmospheres, the  $\Delta$ NLTE model takes as input the stellar parameters and 1D LTE A(Li) and returns  $\delta$ NLTE = 1D NLTE A(Li) - 1D LTE A(Li); whilst the  $A_{\text{REW}}$  model takes stellar parameters and REW and outputs 1D NLTE A(Li). The MAD and RMS errors for these 1000 “interpolation” and 1000 “extrapolation” models are shown in Table 3. All models perform well on average for both interpolation and extrapolation, but extrapolation tends to have more extreme outliers, as seen from the higher RMS for extrapolation. The  $\Delta$ NLTE models have comparable MAD and RMS errors between the 3D leave-one-out cross-validation errors and 1D interpolation and extrapolation errors, with the exceptions of spline and the Cannon. Spline and Kriging for the  $A_{\text{REW}}$  models have more extreme outliers when extrapolating compared

**Table 4.** Error in the predicted abundance correction ( $\Delta\text{NLTE}$ ) and abundance based on equivalent widths ( $A_{\text{REW}}$ ) for verification models (that were not trained on), based on equivalent widths for the 670.8-nm line. The actual values of  $\Delta\text{NLTE}$  and 3D NLTE  $A(\text{Li})$  computed for each verification model is shown in the first row labelled ‘direct synthesis’, while subsequent rows show the errors, as in Table 2.

Line strengths	Method Model	Sun	HD 140283	Procyon	HD 84937
$\Delta\text{NLTE}$	Direct synthesis	−0.065	−0.006	0.040	−0.026
	Spline	0.012	0.021	−0.036	−0.004
	The Cannon	0.000	0.015	−0.045	0.005
	Kriging	0.045	0.016	0.053	−0.001
	SVR	0.007	0.026	−0.029	0.005
	MLP	0.006	0.019	−0.041	0.005
$A_{\text{REW}}$	Direct synthesis	1.100	2.000	1.000	2.200
	Spline	0.000	0.000	−0.004	0.005
	The Cannon	−0.017	−0.006	0.026	0.016
	Kriging	0.011	0.002	−0.005	0.006
	SVR	0.009	−0.008	−0.018	0.008
	MLP	0.002	−0.002	−0.008	0.008

to interpolating; whilst the Cannon, SVR, and MLP have almost comparable RMS, indicating that they perform better extrapolating compared to other methods. Overall, our preferred model, MLP, performs well in both interpolation and extrapolation compared to other models.

Mott et al. (2020) presented polynomial fits with typical errors of 0.020 dex in both  $\Delta\text{NLTE}$  and 3D NLTE REW. We note that although our errors are similar in magnitude, these errors are not directly comparable. This is because the parameters range in the Mott et al. (2020) grid is smaller than ours, and the tests used to derive the error statistics are also different. We believe a leave-one-out cross-validation and an additional test on randomly selected models should be more representative of errors typically seen in real data sets.

In addition, we use MLP to fit the 610.4- and 812.6-nm line. The hyperparameters, leave-one-out cross-validation errors, and verification results are provided in Appendix A.

## 5 DISCUSSION

As shown in Section 3.4, our NLTE results are quantitatively different from previous literature, in particular resulting in significantly more negative abundance corrections. Ordinarily, adjustments to abundance scales like this are tested through the agreement of different abundance diagnostics in standard stars. Due to the very low abundance of lithium, only the resonance line at 670.8 nm is visible in most stars, and indeed even this line is difficult to measure in the Sun where the photospheric abundance is depleted by more than 2 dex relative to the protosolar value (e.g. Asplund et al. 2009). While the subordinate line at 610.4 nm is inaccessible in the solar spectrum, it is sometimes possible to measure in spectra of exceptional quality or when the lithium abundance is strongly enhanced (typically causing the 670.8-nm line to saturate).

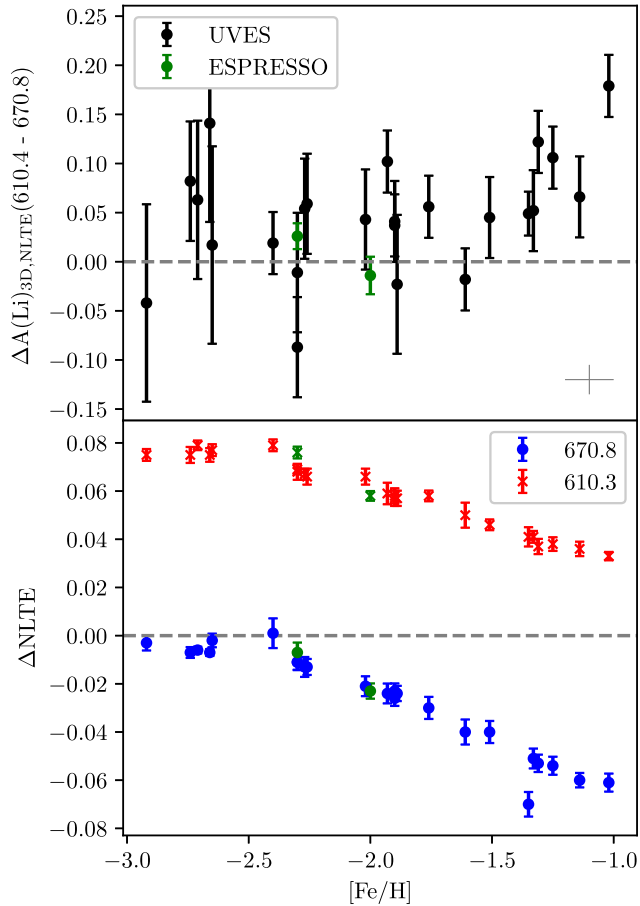
We use a sample of well-studied MSTO stars on the Spite plateau, with equivalent widths measured from high-quality UVES spectra with  $S/N > 400$  (Asplund et al. 2006). We also analyse two benchmark stars with accurately known literature stellar parameters, HD 140283 (Karovicova et al. 2018) and HD 84937 (Casagrande et al. 2011; VandenBerg et al. 2014; Amarsi et al. 2016b), and ESPRESSO spectra with  $S/N \approx 2000$  (Wang et al., in preparation). We find good agreement between abundances measured from the 610.4- and 670.8-nm lines as shown in Fig. 11, but note that the

abundance difference between the two lines is unfortunately rather similar in 3D NLTE and 1D LTE. A small but significant offset is found at  $[\text{Fe}/\text{H}] \approx -1.0$ , that implies an overestimated abundance of the 610.4-nm line relative to the 670.8-nm line. The average abundance difference between the two lines is  $+0.041 \pm 0.055$  dex in 3D NLTE, compared to  $-0.042 \pm 0.053$  in 1D LTE. We emphasize that the 610.4 nm feature has an equivalent width of just 0.2 pm, and the abundance difference could therefore be explained by a minor unrecognized blend with the strength of the order of 0.05 pm at  $[\text{Fe}/\text{H}] = -1$ . Curiously, the Fe II 610.45-nm line is located in the blue wing of the Li I 610.4-nm line and has an equivalent width of approximately 0.3 pm in the solar spectrum. This is in good agreement with the predicted strength (using an oscillator strength from Raassen & Uylings 1998), and should therefore fall very close to expectations in the metal-poor stars. While this blend was not included in the spectrum fits by Asplund et al. (2006) due to its uncertain  $\log gf$  value, we note that they employed a spectrum fitting method that primarily relied on the strength of the red (unblended) wing of the 610.4 nm transition. Assuming the blend did contribute to the measured equivalent widths, the corrected line strengths yield an average abundance difference between the two lines of  $+0.022 \pm 0.049$  dex in 3D NLTE, compared to  $-0.062 \pm 0.048$  dex in 1D LTE.

As our 3D NLTE calculations yield overall lower abundance results than previously, this has important implications for a wide range of science cases. Differences are less significant for e.g. warm and metal-poor stars, where background line opacities in UV are relatively weak. In the measurements presented above, we find for the 22 Spite plateau stars with accurately known temperatures and no obvious Li-enrichment, an average 3D NLTE abundance  $A(\text{Li}) = 2.20 \pm 0.05$  based on the 670.8-nm resonance line, compared to our 1D LTE result  $A(\text{Li}) = 2.23 \pm 0.06$  or the 1D NLTE result of Asplund et al. (2006),  $A(\text{Li}) = 2.21 \pm 0.07$ . This underscores the finding of Asplund et al. (2003) that the 3D NLTE Li abundances in metal-poor halo stars are quite similar to the 1D NLTE and 1D LTE results due to a fortuitous near cancellation of the 3D and NLTE effects.

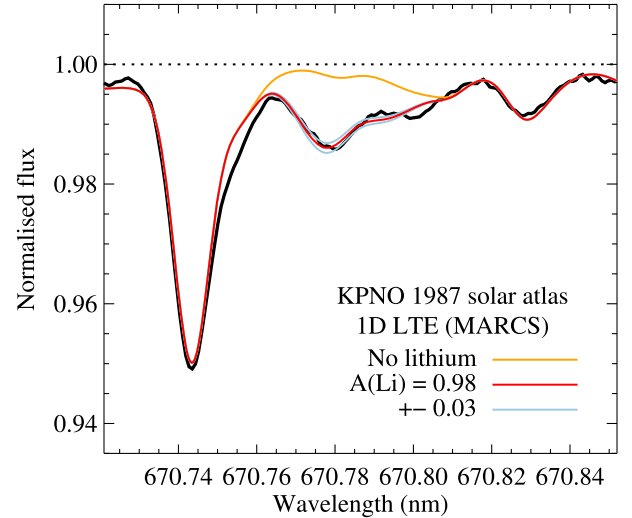
Mucciarelli et al. (2012) performed a commensurate analysis of RGB stars, where their 1D NLTE abundance (with NLTE corrections from Carlsson et al. 1994) measurement  $A(\text{Li}) = 0.97 \pm 0.06$  implies an initial abundance in these field stars of  $A(\text{Li})_0 = 2.28\text{--}2.46$ , where the range represents systematic uncertainties related to the choice





**Figure 11.** Top panel: Comparison between 3D NLTE  $A(\text{Li})$  measurements of the 610.4- and 670.8-nm lines using UVES and ESPRESSO spectra for a sample of Spite plateau stars (see the description in text). A representative systematic error bar due to uncertainties in the stellar parameters is shown in the bottom right-hand corner. Bottom panel: 3D NLTE abundance corrections  $\Delta\text{NLTE}$  (3D NLTE-1D LTE) for the two Li lines.

of temperature scale and stellar evolution models. We revert their  $A(\text{Li})$  to 1D LTE using the abundance corrections from Carlsson et al. (1994), and then apply our own  $\Delta\text{NLTE}$  as described in Sections 3.5 and 4.2; we note that this procedure is not perfect since it depends on the difference in LTE results for the two types of employed 1D model atmospheres (Mucciarelli et al. 2012: Kurucz; here: ATMO) but for our purposes, it is sufficiently accurate given the uncertainties in the predicted Li depletion for RGB stars. Our derived abundances for their stars are significantly lower, at  $A(\text{Li}) = 0.81 \pm 0.06$ , thereby implying initial abundances in the range  $A(\text{Li})_0 = 2.12\text{--}2.30$ , in agreement with the Spite plateau value  $A(\text{Li}) = 2.20 \pm 0.05$  derived above. This unexpectedly good agreement could imply that the surface lithium abundance depletion in metal-poor MSTO stars is mainly caused by destruction rather than non-destructive deposition. We note, however, that theoretical models of mass-dependent depletion through gravitational settling predict a combination of deposition and destruction (Richard et al. 2005) in good agreement with measurements (e.g. Korn et al. 2007; Lind et al. 2008; Meléndez et al. 2010). The initial lithium abundance recovered in these cited studies is, however, still significantly lower than the predicted cosmological abundance level, which opens for additional destruction mechanisms, on the pre-main sequence or otherwise (see Fu et al. 2015).



**Figure 12.** Abundance fit to the solar spectrum in 1D LTE. Our optimum fit has  $\text{EW} = 0.30 \pm 0.02$  pm, which yields a 3D NLTE abundance  $A(\text{Li}) = 0.96 \pm 0.05$  (including systematic errors).

Finally, we have fitted the Hamburg solar flux atlas (Brault & Neckel 1987; Neckel 1999) using the linelist of Ghezzi et al. (2009) as well as of Meléndez et al. (2012) in a 1D LTE analysis with MARCS model atmospheres (Gustafsson et al. 2008). In both cases, our best fit, shown in Fig. 12, the Li line has a line strength of  $\text{EW} = 0.30 \pm 0.02$  pm, which yields a 1D LTE result  $A(\text{Li}) = 0.98 \pm 0.03$  (statistical error bars) in excellent agreement with Harutyunyan et al. (2018). This equivalent width corresponds to a 3D NLTE solar abundance of  $A(\text{Li}) = 0.96 \pm 0.05$  (the error bar here includes indicative systematic errors, representing modelling differences comparing 3D/1D and NLTE/LTE, following Asplund et al. 2009), which is lower than the value  $A(\text{Li}) = 1.05 \pm 0.10$  provided in Asplund et al. (2009). It implies that the solar surface Li abundance is depleted by a factor of 200 relative to the protosolar value as implied by the CI meteoritic abundance  $A(\text{Li}) = 3.26 \pm 0.05$  (Lodders, Palme & Gail 2009, renormalized to  $A(\text{Si}) = 7.51$  from Amarsi & Asplund 2017).

## 6 CONCLUSIONS

In this paper, we presented our 3D NLTE grid, covering the full FGK-type star parameter range, and a broad range in lithium abundances. We make available our grid alongside a set of interpolation routines in the BREIDABLIK package. These interpolation routines can be used to interpolate line profiles to arbitrary stellar parameters using the Kriging technique, achieving a median absolute error of 0.012 dex; we interpolate line strengths using MLP, achieving a median absolute error of 0.012 dex when deriving abundance corrections ( $\Delta\text{NLTE}$ ) and 0.01 dex when deriving the abundance from an REW ( $A_{\text{REW}}$ ). In addition, 1D departure coefficients are made available through Amarsi et al. (2020), and Amarsi (2020).

We have shown the importance of taking into account line blocking in the solution of the statistical equilibrium, and find significant differences with respect to previous works. These lower inferred abundances for all three Li lines may have important implications for a wide range of science cases. As a demonstration, we redetermine the lithium abundances in samples of metal-poor MSTO and RGB stars, and find that our 3D NLTE abundance estimates are lower than previous 1D NLTE estimates by 0.01 and 0.16 dex, respectively; after

correcting for the predicted effects of dilution due to dredge-up, we find that the lithium abundances measured in metal-poor RGB stars agree well with those measured in the MSTO stars.

In future work, we intend to apply our calculations to varying isotopic ratios in metal-poor stars, in order to revisit the cosmological  ${}^6\text{Li}$  problem (e.g. Smith et al. 1998; Asplund et al. 2006; Lind et al. 2013).

## ACKNOWLEDGEMENTS

This research was supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was supported by computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the National Computational Merit Allocation Scheme and the ANU Merit Allocation Scheme (project y89). MA gratefully acknowledges funding from the Australian Research Council (grants FL110100012 and DP150100250). AMA acknowledges support from the Swedish Research Council (VR 2016-03765), and the project grant ‘The New Milky Way’ (KAW 2013.0052) from the Knut and Alice Wallenberg Foundation. KL acknowledges funds from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 852977).

## DATA AVAILABILITY

Our synthetic spectra and interpolation routines are all made publicly available through the BREIDABLIK package, available at <https://github.com/ellawang44/Breidablik>.

## REFERENCES

Amarsi A. M., 2020, Grid/NLTE, Zenodo  
 Amarsi A. M., Asplund M., 2017, *MNRAS*, 464, 264  
 Amarsi A. M., Asplund M., Collet R., Leenaarts J., 2016a, *MNRAS*, 455, 3735  
 Amarsi A. M., Lind K., Asplund M., Barklem P. S., Collet R., 2016b, *MNRAS*, 463, 1518  
 Amarsi A. M., Nordlander T., Barklem P. S., Asplund M., Collet R., Lind K., 2018, *A&A*, 615, A139  
 Amarsi A. M. et al., 2020, *A&A*, 642, A62  
 Andrae R. et al., 2018, *A&A*, 616, A8  
 Anstee S. D., O’Mara B. J., 1995, *MNRAS*, 276, 859  
 Asplund M., Nordlund, Å., Trampedach R., Stein R. F., 1999, *A&A*, 346, L17  
 Asplund M., Nordlund Å., Trampedach R., Allende Prieto C., Stein R. F., 2000, *A&A*, 359, 729  
 Asplund M., Carlsson M., Botnen A. V., 2003, *A&A*, 399, L31  
 Asplund M., Lambert D. L., Nissen P. E., Primas F., Smith V. V., 2006, *ApJ*, 644, 229  
 Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481  
 Barklem P. S., 2016, *A&AR*, 24, 9  
 Barklem P. S., O’Mara B. J., 1997, *MNRAS*, 290, 102  
 Barklem P. S., Belyaev A. K., Asplund M., 2003, *A&A*, 409, L1  
 Basri G., Marcy G. W., Graham J. R., 1996, *ApJ*, 458, 600  
 Baumann P., Ramírez I., Meléndez J., Asplund M., Lind K., 2010, *A&A*, 519, A87  
 Beckmann A., Böklen K. D., Elke D., 1974, *Zeitschrift fur Physik*, 270, 173  
 Belyaev A. K., Barklem P. S., 2003, *Phys. Rev. A*, 68, 062703  
 Bensby T., Lind K., 2018, *A&A*, 615, A151  
 Bildsten L., Brown E. F., Matzner C. D., Ushomirsky G., 1997, *ApJ*, 482, 442

Botnen A., Carlsson M., 1999, in Shoken M. M., Kohji T., Tomoyuki H., eds, *Multi3D, 3D Non-LTE Radiative Transfer*, Kluwer, Boston, MA, p. 379  
 Bouvier J., 2008, *A&A*, 489, L53  
 Brault J. W., Neckel H., 1987, *Spectral Atlas of Solar Absolute Disk-Averaged and Disk-Center Intensity from 3290 to 12510 Å*, Available at: [tp://ftp.hs.uni-hamburg.de/pub/outgoing/FTS-Atlas](http://ftp.hs.uni-hamburg.de/pub/outgoing/FTS-Atlas)  
 Buder S. et al., 2018, *MNRAS*, 478, 4513  
 Carlos M. et al., 2019, *MNRAS*, 485, 4052  
 Carlsson M., 1986, *Uppsala Astron. Obs. Rep.*, 33  
 Carlsson M., Rutten R. J., Bruls J. H. M. J., Shchukina N. G., 1994, *A&A*, 288, 860  
 Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramirez I., Melendez J., Bensby T., Feltzing S., 2011, *A&A*, 530, A138  
 Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q., Gilmore G., 2016, preprint (arXiv:e-print)  
 Castro M., Vauclair S., Richard O., Santos N. C., 2008, *Mem. Soc. Astron. Ital.*, 79, 679  
 Chabrier G., Baraffe I., Plez B., 1996, *ApJ*, 459, L91  
 Chang C.-C., Lin C.-J., 2011, *ACM Trans. Intell. Syst. Technol.*, 2, 27:1  
 Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *ApJ*, 823, 102  
 Collet R., Asplund M., Thévenin F., 2005, *A&A*, 442, 643  
 Collet R., Asplund M., Trampedach R., 2006, *ApJ*, 644, L121  
 Croft H., Dickinson A. S., Gadea F. X., 1999, *MNRAS*, 304, 327  
 Cyburt R. H., Fields B. D., Olive K. A., Yeh T.-H., 2016, *Rev. Mod. Phys.*, 88, 015004  
 Delgado Mena E. et al., 2015, *A&A*, 576, A69  
 Do Nascimento J. D. J., Castro M., Meléndez J., Bazot M., Théado S., Porto de Mello G. F., de Medeiros J. R., 2009, *A&A*, 501, 687  
 Dotter A., 2016, *ApJS*, 222, 8  
 Drawin H.-W., 1968, *Zeitschrift fur Physik*, 211, 404  
 Fabbro S., Venn K. A., O’Brian T., Bialek S., Kieley C. L., Jahandar F., Monty S., 2017, *MNRAS*, 475, 2978  
 Fields B. D., 2011, *Ann. Rev. Nucl. Part. Sci.*, 61, 47  
 Fu X., Bressan A., Molaro P., Marigo P., 2015, *MNRAS*, 452, 3256  
 Geurts P., Ernst D., Wehenkel L., 2006, *Mach. Learn.*, 63, 3  
 Ghezzi L., Cunha K., Smith V. V., Margheim S., Schuler S., de Araújo F. X., de la Reza R., 2009, *ApJ*, 698, 451  
 Gonzalez G., 2008, *MNRAS*, 386, 928  
 Gonzalez G., 2015, *MNRAS*, 446, 1020  
 Gruyters P., Nordlander T., Korn A. J., 2014, *A&A*, 567, A72  
 Gruyters P. et al., 2016, *A&A*, 589, A61  
 Gustafsson B., Edvardsson B., Eriksson K., Jørgensen U. G., Nordlund, Å., Plez B., 2008, *A&A*, 486, 951  
 Harutyunyan G., Steffen M., Mott A., Caffau E., Israelian G., González Hernández J. I., Strassmeier K. G., 2018, *A&A*, 618, A16  
 Hinton G. E., 1990, *Machine learning*. Morgan Kaufmann Elsevier, U. S. p. 555  
 Ho A. Y. Q., Ness M., Hogg D. W., Rix H.-W., 2016, *Astrophysics Source Code Library*, record ascl:1602.010  
 Ho A. Y. Q. et al., 2017, *ApJ*, 836, 5  
 Israelian G. et al., 2009, *Nature*, 462, 189  
 James G., Witten D., Hastie T., Tibshirani R., 2013, *An Introduction to Statistical Learning*. Springer-Verlag, Berlin  
 Jeffries R. D., 2014, in Lebreton Y., Valls-Gabaud D., Charbonnel D., eds, *EAS Publications Series*, Vol. 65, *The Ages of Stars*, Cambridge Univ. Press, Cambridge, p. 289  
 Karovicova I. et al., 2018, *MNRAS*, 475, L81  
 Kaulakys B., 1985, *J. Phys. B Atom. Mol. Phys.*, 18, L167  
 Kaulakys B., 1991, *J. Phys. B Atom. Mol. Phys.*, 24, L127  
 Kiselman D., 1997, *ApJ*, 489, L107  
 Kiselman D., 1998, *A&A*, 333, 732  
 Klevas J., Kučinskas A., Steffen M., Caffau E., Ludwig H. G., 2016, *A&A*, 586, A156  
 Klevas J., Kučinskas A., Wedemeyer S., Ludwig H. G., 2018, *Contrib. Astron. Obs. Skalnaté Pleso*, 48, 280

- Korn A. J., Grundahl F., Richard O., Mashonkina L., Barklem P. S., Collet R., Gustafsson B., Piskunov N., 2007, *ApJ*, 671, 402
- Kuhn M., Johnson K., 2013, *Applied Predictive Modelling*. Springer, New York
- Leenaarts J., Carlsson M., 2009, MULTI3D: A Domain-Decomposed 3D Radiative Transfer Code, ASP Conf. Ser., Astronomical Society of the Pacific, San Francisco, p. 87
- Leung H. W., Bovy J., 2019, *MNRAS*, 483, 3255
- Lind K., Korn A. J., Barklem P. S., Grundahl F., 2008, *A&A*, 490, 777
- Lind K., Asplund M., Barklem P. S., 2009, *A&A*, 503, 541
- Lind K., Melendez J., Asplund M., Collet R., Magic Z., 2013, *A&A*, 554, A96
- Lindgård A., Nielson S. E., 1977, *At. Data Nucl. Data Tables*, 19, 533
- Lodders K., Palme H., Gail H.-P., 2009, *Landolt Börnstein*, 4B, 44
- Magic Z. et al., 2013, *A&A*, 557, A26
- Meléndez J. et al., 2010, *Ap&SS*, 328, 193
- Meléndez J., Casagrande L., Ramírez I., Asplund M., Schuster W. J., 2010, *A&A*, 515, L3
- Meléndez J. et al., 2012, *A&A*, 543, A29
- Mott A., Steffen M., Caffau E., Spada F., Strassmeier K. G., 2017, *A&A*, 604, A44
- Mott A., Steffen M., Caffau E., Strassmeier K. G., 2020, *A&A*, 638, A58
- Mucciarelli A., Salaris M., Bonifacio P., 2012, *MNRAS*, 419, 2195
- Mucciarelli A., Salaris M., Bonifacio P., Monaco L., Villanova S., 2014, *MNRAS*, 444, 1812
- Neckel H., 1999, *Sol. Phys.*, 184, 421
- Ness M., Hogg D. W., Rix H.-W., Ho A. Y. Q., Zasowski G., 2015, *ApJ*, 808, 16
- Nordlander T., Lind K., 2017, *A&A*, 607, A75
- Nordlander T., Korn A. J., Richard O., Lind K., 2012, *ApJ*, 753, 48
- Nordlander T., Amarsi A. M., Lind K., Asplund M., Barklem P. S., Casey A. R., Collet R., Leenaarts J., 2017, *A&A*, 597, A6
- Osorio Y., Barklem P. S., Lind K., Asplund M., 2011, *A&A*, 529, A31
- Pavlenko Y. V., Magazzu A., 1996, *A&A*, 311, 961
- Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, *ApJS*, 192, 3
- Paxton B. et al., 2013, *ApJS*, 208, 4
- Paxton B. et al., 2015, *ApJS*, 220, 15
- Peach G., Saraph H. E., Seaton M. J., 1988, *J. Phys. B Atom. Mol. Phys.*, 21, 3669
- Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
- Pitrou C., Coc A., Uzan J.-P., Vangioni E., 2018, *Phys. Rep.*, 754, 1
- Prantzos N., 2012, *A&A*, 542, A67
- Prantzos N., de Laverny P., Guiglion G., Recio-Blanco A., Worley C. C., 2017, *A&A*, 606, A132
- Puchalski M., Pachucki K., 2009, *Phys. Rev. A*, 79, 032510
- Raassen A. J. J., Uylings P. H. M., 1998, *A&A*, 340, 300
- Ramírez I., Fish J. R., Lambert D. L., Allende Prieto C., 2012, *ApJ*, 756, 46
- Richard O., Michaud G., Richer J., 2005, *ApJ*, 619, 538
- Ryan S. G., 2000, *MNRAS*, 316, L35
- Sansonetti C. J., Simien C. E., Gillaspay J. D., Tan J. N., Brewer S. M., Brown R. C., Wu S., Porto J. V., 2011, *Phys. Rev. Lett.*, 107, 023001
- Sbordone L. et al., 2010, *A&A*, 522, A26
- Schönrich R., Binney J., 2009, *MNRAS*, 396, 203
- Sestito P., Randich S., 2005, *A&A*, 442, 615
- Smith V. V., Lambert D. L., Nissen P. E., 1998, *ApJ*, 506, 405
- Snider S., Allende Prieto C., von Hippel T., Beers T. C., Sneden C., Qu Y., Rossi S., 2001, *ApJ*, 562, 528
- Soderblom D. R., 2010, *Annu. Rev. A&A*, 48, 581
- Spite F., Spite M., 1982, *A&A*, 115, 357
- Steenbock W., Holweger H., 1984, *A&A*, 130, 319
- Takeda Y., 2019, *A&A*, 622, A107
- Takeda Y., Kawonomoto S., 2005, *PASJ*, 57, 45
- Ting Y.-S., Conroy C., Rix H.-W., Cargile P., 2019, *ApJ*, 879, 69
- Uitenbroek H., 1998, *ApJ*, 498, 427
- Ushomirsky G., Matzner C. D., Brown E. F., Bildsten L., Hilliard V. G., Schroeder P. C., 1998, *ApJ*, 497, 253
- VandenBerg D. A., Bond H. E., Nelan E. P., Nissen P. E., Schaefer G. H., Harmer D., 2014, *ApJ*, 792, 110
- Vapnik V. N., 1995, *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, NY
- Yan Z.-C., Tambasco M., Drake G. W. F., 1998, *Phys. Rev. A*, 57, 1652
- Žerjal M. et al., 2019, *MNRAS*, 484, 4591

## APPENDIX A: INTERPOLATION OF THE 610.4- AND 812.6-NM LINES

In addition to the line strength models for the 670.8-nm line presented in the main text, we also developed line strength interpolation models for both the 610.4- and 812.6-nm lines. In Section 4.2, we found that the best performing model is MLP for both  $\Delta$ NLTE and  $A_{\text{REW}}$ , therefore, we use MLP for these new models. In addition, we use the same five-fold hyperparameter tuning and verification models as previously established to develop these new models.

Table A1 tabulates the best hyperparameters found through five-fold cross-validation. The  $\Delta$ NLTE hyperparameters are similar to the model presented in Section 4.2.2 whilst the  $A_{\text{REW}}$  hyperparameters differ. This is likely due to the dynamic range and shape of the  $A_{\text{REW}}$  data set changing more compared to the  $\Delta$ NLTE data set when changing Li lines.

The leave-one-out cross-validation fitting statistics is shown in Table A2. We find that the errors are smaller than the models presented in Section 4.2 because the dynamic range of  $\Delta$ NLTE and  $A_{\text{REW}}$  for the 610.4 and 812.6-nm lines are smaller than the dynamic range of the 670.8-nm line. The training and execution time for these models are comparable to MLP times in Table 3.

The verification results are shown in Table A3. Procyon for the 810.6-nm line is not included as it is immeasurably weak. Overall, for both  $\Delta$ NLTE and  $A_{\text{REW}}$ , most errors in verification models fluctuate around the MAD leave-one-out cross-validation errors. Therefore, both the  $\Delta$ NLTE and  $A_{\text{REW}}$  final models for the 610.4- and 810.6-nm line are consistent with their leave-one-out cross-validation errors.

**Table A1.** Optimized MLP hyperparameters for  $\Delta$ NLTE and  $A_{\text{REW}}$  for both the 610.4- and 810.6-nm line, based on five-fold cross-validation.  $n_l$  is the number of layers,  $n$  is the number of neurons per layer,  $\alpha$  is the L2 parameter, and  $f_{\text{act}}$  is the activation function.

Line strengths	Li line (nm)	$n_l$	$n$	$\alpha$	$f_{\text{act}}$
$\Delta$ NLTE	610.4	2	900	$10^{-5}$	ReLU
	812.6	2	850	$10^{-1}$	ReLU
$A_{\text{REW}}$	610.4	2	750	$10^{-9}$	ReLU
	812.6	2	750	$10^{-2}$	ReLU

**Table A2.** Interpolation errors using an MLP model to predict  $\Delta$ NLTE and  $A_{\text{REW}}$  for both the 610.4- and 810.6-nm line, based on an average over leave-one-out cross-validation errors using the 3D grid.

Line strengths	Li line (nm)	MAD	RMS
$\Delta$ NLTE	610.4	0.007	0.010
	812.6	0.007	0.011
$A_{\text{REW}}$	610.4	0.006	0.008
	812.6	0.007	0.009



**Table A3.** Error in predicted  $\Delta\text{NLTE}$  and  $A_{\text{REW}}$  using MLP line strength interpolation models trained on the 610.4 nm and 810.6 lines on verification models (that were not trained on). The actual  $\Delta\text{NLTE}$  and  $A_{\text{REW}}$  are shown in the direct synthesis. The verification models were tailored to specific stars, indicated by name. The 810.6-nm line is neglected for Procyon as it is immeasurably weak.

Line strengths	Li line (nm)	Method Model	Sun	HD 140283	Procyon	HD 84937
$\Delta\text{NLTE}$	610.4	Direct synthesis	0.021	0.069	0.080	0.057
		MLP	0.012	0.016	−0.022	0.002
	810.6	Direct synthesis	0.001	0.060		0.039
		MLP	0.009	0.015		0.011
$A_{\text{REW}}$	610.4	Direct synthesis	1.100	2.000	1.000	2.200
		MLP	0.003	−0.002	0.001	0.003
	810.6	Direct synthesis	1.100	2.000		2.200
		MLP	−0.004	0.004		0.009

This paper has been typeset from a  $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  file prepared by the author.