The STAGGER-grid: A Grid of 3D Stellar Atmosphere Models

II. Horizontal and Temporal Averaging and Spectral Line Formation

Z. Magic[1,2], R. Collet[2,1], W. Hayek[1,2] and M. Asplund[2,1]

[1] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
e-mail: magic@mpa-garching.mpg.de
[2] Research School of Astronomy & Astrophysics, Australian National University, Cotter Road, Weston ACT 2611, Australia

ABSTRACT

Aims. We study the implications of averaging methods with different reference depth scales for 3D hydrodynamical model atmospheres computed with the STAGGER-code. The temporally and spatially averaged (hereafter denoted as ⟨3D⟩) models are explored in the light of local thermodynamic equilibrium (LTE) spectral line formation by comparing spectrum calculations using full 3D atmosphere structures with those from ⟨3D⟩ averages.
Methods. We explore methods for computing mean ⟨3D⟩ stratifications from the STAGGER-grid time-dependent 3D radiative hydrodynamical atmosphere models by considering four different reference depth scales (geometrical depth, column-mass density, and two optical depth scales). Furthermore, we investigate the influence of alternative averages (logarithmic or enforced hydrostatic equilibrium, flux-weighted temperatures). For the line formation we compute curves of growth for Fe I and Fe II lines in LTE .
Results. The resulting ⟨3D⟩ stratifications for the four reference depth scales can be considerably different. We find typically that in the upper atmosphere and in the superadiabatic region just below the optical surface, where the temperature and density fluctuations are highest, the differences become considerable and increase for higher T_{eff}, lower $\log g$, and lower [Fe/H]. The differential comparison of spectral line formation shows distinctive differences depending on which ⟨3D⟩ model is applied. The averages over layers of constant column-mass density yield the best mean ⟨3D⟩ representation for LTE line formation, while the averages on layers at constant geometrical height are the least appropriate. Unexpectedly, the usually preferred averages over layers of constant optical depth are prone to the increasing interference of the reversed granulation towards higher effective temperature, in particular at low metallicity.

Key words. convection – hydrodynamics – radiative transfer – line: formation – stars: abundances – stars: atmospheres – stars: fundamental parameters– stars: general– stars: late-type – stars: solar-type

1. Introduction

Theoretical model atmospheres are necessary in order to interpret stellar fluxes and derive individual characteristics of stars, like stellar parameters and chemical abundances. In recent decades, successive improvements of the often used one-dimensional (1D) hydrostatic atmosphere models have confirmed their predictive capabilities (see e.g. Gustafsson et al. 2008) but also highlighted their limitations. In fact, these 1D models make use of several simplifications in favor of computational ease, the most prominent one being the treatment of convection with the mixing-length theory (MLT, Böhm-Vitense 1958; Henyey et al. 1965). The latter entails several free parameters, in particular the free mixing-length parameter, α_{MLT}, which is a priori unknown, hence is normally calibrated for the Sun by observations and assumed constant for all stars. Moreover, the calculation of synthetic spectral absorption lines in 1D requires the additional calibration of micro- and macro-turbulence parameters (ξ_{turb} and χ_{turb}, respectively) in order to account properly for the contribution of non-thermal convective and turbulent motions to the broadening of spectral line profiles.

Most of the limitations of 1D modeling of convection can be overcome only by performing time-dependent, three dimensional (3D), radiative-hydrodynamical (RHD) calculations (see Nordlund et al. 2009, and references therein). The goal of 3D simulations is to provide realistic ab initio models where stellar surface convection emerges self-consistently from first principles. Compared to 1D models, such 3D RHD models are able, for the Sun in particular, to predict additional observable features of stars associated with stellar surface velocity fields and temperature and density inhomogeneities, e.g. surface granulation pattern, line asymmetries, and center-to-limb variation (CLV; e.g. Asplund et al. 2000b; Pereira et al. 2013). In order to systematically study such properties of stars with a realistic approach, we have computed a large grid of 3D models using the STAGGER-code, covering a wide range in stellar parameters[1] (T_{eff}, $\log g$, and [Fe/H]) for late-type (spectral type FGK) stars (see Magic et al. 2013, hereafter Paper I).

It is advantageous to reduce the relatively large amount of data from the full 3D atmospheric models to temporally and spatially averaged (hereafter ⟨3D⟩) representations. However, this reduction is based at the expense of physical self-consistency (see Atroshchenko & Gadun 1994). Nonetheless, in this way one can deal with more manageable atmospheric data structures

---

[1] In the following, we refer always to stellar *atmospheric* parameters.

compared to the otherwise enormous amount of information associated with the full 3D models. These mean ⟨3D⟩ stratifications are usually compared with classical 1D hydrostatic atmosphere models. Nordlund & Stein (2001) pointed out that the large-amplitude fluctuations in the superadiabatic region leads to deviations from the hydrostatic equilibrium. Furthermore, the 3D data sets incorporate quantities emerging from the hydrodynamics and associated with convection itself, e.g. self-consistent velocity fields and turbulent pressure, for which there are no physically consistent counterparts in the case of 1D hydrostatic models.

The definition of the ⟨3D⟩ stratifications is not unambiguous nor unique, but depends largely on the choice of reference depth scale. When dealing with the analysis of the atmospheric layers above the optical surface, monochromatic or Rosseland optical depth scales are usually considered the appropriate choice since these are the natural reference depth scales that are used to describe radiative transfer processes in the photosphere. On the other hand, the optical depth loses somewhat its usefulness in the very deep optically thick layers below the optical surface. There, other reference scales are best suited to describe the main properties of the stellar stratification. Also, the bimodal and highly asymmetric distribution of bulk upflows and downflows in the convective zone complicates the definition of a meaningful unique average value, particularly near the surface, at the transition between convectively unstable and stable regions.

Uitenbroek & Criscuoli (2011) investigated the application of ⟨3D⟩ models to spectral line formation. They computed and compared continuum and atomic line intensities and their respective CLV from ⟨3D⟩ and 3D models. They concluded that a mean ⟨3D⟩ stratification is insufficient to represent the full 3D atmosphere model in the light of spectral analysis. As reasons for the latter they list the nonlinearity of the Planck function, formation of molecules and the asymmetry of convective motions.

The present work constitutes the second paper in the STAGGER-grid series. Here, we want to explore the following key question: what is the most meaningful ⟨3D⟩ stratification derived from the full 3D data set? We discuss this question in particular in the context of radiative transfer applications such as spectral line formation. We investigate which averaging method leads to the best choice for reproducing the spectral line absorption features by probing the latter with fictitious Fe I and Fe II lines with different strengths and excitation potentials.

2. Averaging 3D models

2.1. The 3D atmosphere models

The 3D models that form the basis of the present work were computed with the STAGGER-code; for a general description of our grid of 3D models, we refer the reader to Paper I. In short, the STAGGER-code solves the time-dependent, 3D hydrodynamical equations coupled with realistic non-gray radiative transfer. We utilize an updated version of the realistic state-of-the-art equation of state (EOS) by Mihalas et al. (1988). Continuum and sampled line opacity are taken primarily from the MARCS package (Gustafsson et al. 2008; see also references in Paper I). The radiative transfer is solved for nine angles along long characteristics with a slightly modified version of the Feautrier (1964) method. The opacity-binning method with 12 opacity bins is

applied to all STAGGER-grid models to reduce the computational burden while still accounting for non-grey radiative transfer (see Nordlund 1982; Skartlien 2000). Our simulations are of the so-called *box-in-a-star* type and cover only a small representative volume of stellar surface typically including about ten granules horizontally and spanning about 14 pressure scale heights vertically. The numerical resolution of the Cartesian grid is 240^3. It features a non-equidistant vertical axis in order to enhance resolution in the layers with the steepest temperature gradients. The vertical boundaries are open, while the horizontal ones are periodic.

2.2. Computing temporal and horizontal averages

We computed various temporal and horizontal averages for a large number of physical quantities of interest. For the spatial (horizontal) averages, we compute ⟨3D⟩ stratifications by considering four different reference depth scales and averaging the various physical quantities on layers of constant

- geometrical height, z,
- column mass density, $m = \int \rho \, dz$,
- Rosseland optical depth, $\tau_{\mathrm{Ross}} = \int (\rho \kappa_{\mathrm{Ross}}) \, dz$,
- optical depth at 500 nm, $\tau_{500} = \int (\rho \kappa_{500}) \, dz$

(hereafter denoted by $\langle 3D \rangle_z$, $\langle 3D \rangle_m$, $\langle 3D \rangle_{\mathrm{Ross}}$ and $\langle 3D \rangle_{500}$, respectively), where ρ is the gas density, and κ_{Ross} and κ_{500} are the Rosseland mean opacity and opacity at 500 nm, respectively, both defined as cross-sections per unit mass. The geometrical averages $\langle 3D \rangle_z$ are easily taken directly from the output of the STAGGER-code, since the numerical mesh of this code is of Eulerian nature. For the other three (Lagrangian-like) averages, the original data sets have to be remapped to their respective new reference depth scale by individually interpolating each column of each 3D simulation snapshot (see 2.2.2). Furthermore, we also considered four additional averages:

- flux-weighted average temperature, $\langle T^4 \rangle$,
- average brightness temperature at 500nm, $\langle T_{\mathrm{rad}} \rangle$,
- logarithmic average, $\langle 3D \rangle_{\mathrm{log}}$,
- enforced-hydrostatic-equilibrium average, $\langle 3D \rangle_{\mathrm{HSE}}$.

We determine the flux-weighted temperature stratification $\langle T^4 \rangle$ by evaluating the spatial averages of T^4, motivated by the Stefan-Boltzmann law for wavelength-integrated radiative flux. The brightness temperature average T_{rad} is computed using the expression $B_{500}^{-1}(\langle B_{500}(T) \rangle)$, where B_{500} and B_{500}^{-1} denote the Planck function at 500 nm and its inverse, respectively (see also Sect. 3.1); the depth-dependent $\langle T_{\mathrm{rad}} \rangle$ is thus to be interpreted as the equivalent brightness temperature corresponding to the average black-body emission at 500 nm from each layer. For $\langle 3D \rangle_{\mathrm{log}}$ we define spatial averages of a given 3D variable X as $\exp(\langle \log X \rangle)$. Finally, since the ⟨3D⟩ models do not fulfill in general the hydrostatic equilibrium condition (see App. A.2), for the $\langle 3D \rangle_{\mathrm{HSE}}$ averages we *enforce* hydrostatic equilibrium by adjusting the density and adjusting the thermodynamic pressure p_{th} consistently with the EOS, until hydrostatic equilibrium is attained. We emphasize that the proper enforcement of hydrostatic equilibrium requires that one considers both the thermodynamic p_{th} and turbulent p_{turb} contributions to total pressure p_{tot}: the gas pressure in the atmosphere is in fact significantly reduced because of the structural support provided by turbulent pressure. Then, a new geometrical depth z is computed (see Eq. A.2).

Classical hydrostatic 1D models of stellar atmospheres are often defined and computed on an optical depth scale, since this allows to easily adjust the numerical resolution where it is most necessary to achieve the highest accuracy in the solution of the radiative transfer equation in the atmospheric layers, both during the modeling itself and during line formation calculations. Therefore, especially for radiative transfer-oriented applications, these 1D models can be compared most naturally with averages of corresponding 3D models on constant optical depth, $\langle 3D \rangle_{\mathrm{Ross}}$ or $\langle 3D \rangle_{500}$. In Paper I, in particular, we adopted $\langle 3D \rangle_{\mathrm{Ross}}$ as our standard averaging choice; one of the main reasons why we chose $\langle 3D \rangle_{\mathrm{Ross}}$ over $\langle 3D \rangle_{500}$ is that during the scaling of the simulations and the construction of the initial snapshots, the top physical boundary of essentially all models reached up to $\langle \log \tau_{\mathrm{Ross}} \rangle_{\mathrm{top}} \approx -6.0$ (see Paper I). On the contrary, the vertical extent of the simulations in terms of optical depth at 500 nm varies depending on stellar parameters ($\log g$ in particular) owing to the concomitant variations of the opacity at 500 nm as a function of temperature and density. Therefore, the $\langle 3D \rangle_{500}$ models require in general a careful extrapolation at the top to be extended up to $\log \tau_{500} \approx -6.0$ (see Sect. 2.2.3). On the other hand, while regarding the conditions below surface it is better to use $\langle 3D \rangle_z$ or $\langle 3D \rangle_m$.

While $\langle 3D \rangle_{\mathrm{Ross}}$ or $\langle 3D \rangle_{500}$ represent natural reference depth scales for the mean photospheric stratification, $\langle 3D \rangle_z$ or $\langle 3D \rangle_m$ are better suited for the description of the average physical conditions below the stellar surface, e.g. only the geometrical averages are fulfilling conservation of momentum and energy (see App. A.2).

In late-type stellar atmospheres, the continuum opacity κ_λ in the optical is dominated by the H^- bound-free absorption which depends sensitively on temperature ($\sim T^{10}$). Therefore, even small fluctuations in T will result in large variations in κ_λ, which in turn will lead to a high degree of spatial corrugation of layers at constant optical depth (see Stein & Nordlund (1998)). Furthermore, owing to such highly non-linear behavior of the H^- opacity, temperature fluctuations around the average will be reduced by interpolation to layers of constant of optical depth (see Sect. 4.1).

We note briefly that only the geometrical averages $\langle 3D \rangle_z$ preserve the general conservation properties of the basic hydrodynamical equations. Therefore, only the $\langle 3D \rangle_z$ stratifications appear sufficiently close to hydrostatic equilibrium on average and fulfill energy conservation conditions. Because of properties of the different types of $\langle 3D \rangle$ models vary significantly among the different averaging methods, the best choice depends strongly on the intended particular application. It is therefore very important to use $\langle 3D \rangle$ models with care (see App. A).

2.2.1. Basic averaging procedure

We proceeded with the following steps in order to obtain the $\langle 3D \rangle$ models:

1. Retrieval of 3D variables of interest;
2. Interpolation to new reference depth scale;
3. Computation of horizontal averages and statistics;
4. Extrapolation of horizontal averages, if necessary;
5. Computation of temporal averages.

In case of the geometrical averages $\langle 3D \rangle_z$, steps 2 and 4 are unnecessary and are therefore skipped. Owing to the generally non-linear response of the various physical quantities as a function of basic independent variables and the EOS, the interpolation to a new reference depth scale should be performed after the retrieval of the variables. In particular, because of these non-linearities, we caution against the derivation of thermodynamic variables via the EOS by utilizing averaged independent variables interpolated to the new reference depth scale, since the spatial averaging will inevitably break the physical self-consistency present in the full original 3D data (see Sect. 2.2.2 and A.3).

At the vertical boundaries of our simulation box are so called *ghost-zones*, each consisting of five layers at the top and bottom. Their sole purpose is to numerically define the boundary conditions at both vertical ends; they do not contain physically meaningful values, hence we excluded them before the averaging procedure.

To speed up the calculations without noticeably degrading the statistical properties, when computing the averages we consider only every fourth column of the 3D data cubes in both horizontal directions (x and y), which means that the initial $N_x N_y = 240^2$ columns are reduced down to 60^2. The vertical extent of the columns is unchanged with $N_z = 230$ (geometrical) or 101 (all other reference depth scales). Tests ensured that this horizontal reduction does not influence the horizontal averages due to the still large sample of vertical columns considered, and the multiple snapshots included in the temporal averaging.

For step 3, we used an arithmetic mean to compute the average values of variable X for snapshot t at each horizontal layer z:

$$\langle X \rangle_{z,t} = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} X_{xyz,t} \tag{1}$$

with N_x and N_y being the number of horizontal elements. For exponentially varying variables like density and pressure we computed also logarithmic averages, i.e., replacing X_{xyz} with $\log X_{xyz}$ in Eq. 1, denoting the models with $\langle 3D \rangle_{\mathrm{log}}$. In the final step 5, temporal averages are evaluated with

$$\langle X \rangle_z = \frac{1}{N_t} \sum_{t=1}^{N_t} \langle X \rangle_{z,t} \tag{2}$$

with $N_t \approx 100 - 150$ being the total number of snapshots considered for each simulation. In the present work, the combined temporal and spatial averages of variable X are always denoted with $\langle X \rangle_{\tilde{z}}$, where $\tilde{z}$ is the considered reference depth scale.

Since the 3D structures display a great plethora of details, for each relevant 3D variable we also determine a number of additional statistical properties (standard deviation σ, root mean square, minimum-maximum range, and histograms of the distribution of values) at each horizontal layer, which are presented and discussed in Sect. 4. As for the spatial averages, the standard deviation and the root mean square are evaluated in step 3 for each layer z using the same basic expression as in Eq. 1 and, if necessary, doubly extrapolated at the top as in step 2 and 4 (see Sect. 2.2.3). Finally, their temporal averages are computed in step 5.

Histograms of the distribution of values are determined separately; we use temporal averages of the depth-dependent extrema of variable X, $\langle \min X \rangle_z$ and $\langle \max X \rangle_z$ to define a depth-dependent range $r_z = [\langle \min X \rangle_z, \langle \max X \rangle_z]$ for the histograms.

For the 3D variable X at time t, we determine a set of 1D histograms, $p_{r,z,t}(X)$, for each individual layer z. The depth-dependent range r_z is resolved with $N_r = 200$ equidistant points; temporal averages $p_{r,z}(X)$ of the histograms are computed using a subset of $N_t = 20$ equidistant snapshots (see Sect. 4.3 for details).

Finally, we also compute averages and associated statistical properties separately for up- and downflows, which we discriminate based on the sign of the vertical component of the velocity. Of course, when computing such averages and statistics, one has to account for the correct filling factor in either case, i.e. for the number of elements $N_{x,y}$ belonging to up- or downflows, respectively (Sect. 4.2).

2.2.2. Interpolation to the new reference depth scale

In order to interpolate to the new reference depth scale (hereafter denoted as $\tilde{z}$) in step 2, we defined for both optical depth scales τ_{Ross} and τ_{500}, a new equidistant logarithmic reference optical depth scale, $\tilde{z} = \log \tilde{\tau}$, from $-5.0, \dots, +5.0$ in steps of 0.1. In the case of averaging based on the column-mass density scale m, we used for the new reference depth scale the column-mass density $\tilde{m}$ normalized to the mean value of m at the optical surface, i.e. $\tilde{z} = \log(\tilde{m}) = \log(m/\langle m \rangle_{\mathrm{surf}})$, where $\langle m \rangle_{\mathrm{surf}}$ is determined at $\langle \tau_{\mathrm{Ross}} = 0 \rangle$, and considered a fixed range from $-3.0, \dots, +2.0$ in steps of 0.05 for all simulations. All variables, X, are remapped column-wise from the original geometrical depth scale to the new reference depth scale, namely $X_{xy}(z) \rightarrow \tilde{X}_{xy}(\tilde{z})$. We used linear interpolation, since quadratic interpolation introduced numerical artifacts in some $\langle 3D \rangle$ models.

We note that due to the remapping to a new reference depth scale, points at a constant optical depth or column-mass density will end up probing and spanning a range of geometrical depths, implying that the averages (and statistical properties) with respect to the new reference depth scale will be qualitatively and quantitatively different from plain horizontal averages on constant geometrical depth (see App. A).

2.2.3. Extrapolation at the top

The vast majority of STAGGER-grid models are sufficiently extended vertically, in particular at the top to embrace the full range of $\log \tilde{\tau}$ with $[-5.0, +5.0]$. The condition $\langle \log \tau_{\mathrm{Ross}} \rangle_{\mathrm{top}} \leq -6.0$, is usually fulfilled for all but a few models. More specifically, surfaces of constant optical depth can become quite corrugated at the top for some giant models and fall outside the physical domain of the simulations, i.e. occasionally one can have $\log \tau_{\mathrm{Ross}}^{\mathrm{top}} > -5.0$ for a limited number of columns. These particular columns are therefore linearly extrapolated to $\log \tau_{\mathrm{Ross}} = -5.0$ to allow the calculation of average quantities in the desired range of optical depths. The extrapolation is needed only for a few giant models ($\log g \leq 2.5$), and the concerned columns are usually only a small fraction ($\lesssim 0.3\%$). Therefore, we regard these extrapolations as negligible in the case of the optical depth scale τ_{Ross}.

For the optical depth scale τ_{500}, the situation is slightly different. The mean optical depth at 500 nm at the top $\langle \log \tau_{500} \rangle_{\mathrm{top}}$ deviates increasingly towards giant models from $\langle \log \tau_{\mathrm{Ross}} \rangle_{\mathrm{top}}$, so that $\langle \log \tau_{500} \rangle_{\mathrm{top}} > -5.0$. Therefore, the necessary extrapolation at the top are considerable, in particular for giant models.

We remark that careless column-wise extrapolation at the top can lead to a largely uncertain and erroneous stratification, which would have a negative impact on spectral line formation. For instance, a wrong density stratification at the top can dramatically affect the ionization balance. In order to limit these extrapolation errors, we first restrict the column-wise extrapolation to the region $\log \tilde{\tau}_{500} \geq \log \tilde{\tau}_{\mathrm{top}}$ where the value $\log \tilde{\tau}_{\mathrm{top}} > -5.0$ is chosen so that only 20% of the columns would require extrapolation up to that level. We then compute the horizontal averages (step 3) and, after that, linearly extrapolate the $\langle 3D \rangle$ models a second time to the original $\log \tilde{\tau}_{\mathrm{top}} = -5.0$ for each time snapshot. This particular extrapolation procedure produces more plausible stratifications since the horizontal $\langle 3D \rangle$ averages exhibit a rather smooth and monotonic behavior with depth at the top compared to individual columns of the 3D data set. Test calculations of data sets, which were truncated at top, revealed the reliability of this *double extrapolation* approach.

Nonetheless, we favor the use of averages on mean Rosseland optical depth, i.e. $\langle 3D \rangle_{\mathrm{Ross}}$ rather than $\langle 3D \rangle_{500}$, since these averages are not plagued by such extrapolation uncertainties. For the extrapolated models on τ_{500}, we kept track of the extent of the applied extrapolation. Hence, the $\langle 3D \rangle_{500}$ averages can be reduced to the extrapolation-free regime at the top afterwards.

3. Comparison of the averaging methods

In the following, we systematically compare the different types of averaging procedures explained in Sec. 2 over a broad range of stellar parameters relative to Rosseland optical depth, i.e. $\langle 3D \rangle_{\tilde{z}} - \langle 3D \rangle_{\mathrm{Ross}}$. For the sake of clarity, we illustrate the properties of average stratifications only for a representative selection of STAGGER-grid models comprising dwarfs and giants ($\log g = 4.5$ and 2.0) at solar and sub-solar metallicity ($[\mathrm{Fe/H}] = 0.0$ and -3.0). Besides the most important thermodynamic state variables, temperature and density, we also investigate averages of electron number density, an important quantity for, e.g., calculations of ionization balance and spectral line formation.

Due to the lack of a unique common global depth scale that is invariant between different averaging methods, we display their results jointly on the averaged Rosseland optical depth scale, $\langle \tau_{\mathrm{Ross}} \rangle$, in order to enable a direct comparison.

3.1. Temperature

We find that the temperature stratifications of the two optical reference depth scales, $\langle 3D \rangle_{\mathrm{Ross}}$ and $\langle 3D \rangle_{500}$, therefore we refrain to show these. Only at the top of the metal-poor stars the $\langle 3D \rangle_{500}$-averages appear cooler ($\sim 5\%$, i.e by $\gtrsim 250\,\mathrm{K}$ at $T_{\mathrm{eff}} = 6000\,\mathrm{K}$). On the other hand, the geometrical $\langle 3D \rangle_z$ and column mass density $\langle 3D \rangle_m$ averages deviate distinctively from the $\langle 3D \rangle_{\mathrm{Ross}}$-stratification (see Fig. 1). In the regime $-1.0 < \log \tau_{\mathrm{Ross}} < 3.0$, both $\langle 3D \rangle_z$ and $\langle 3D \rangle_m$ are cooler by $\sim 5 - 10\%$. At the surface ($\tau_{\mathrm{Ross}} = 0$) the geometrical averages deviate considerably, while the $\langle 3D \rangle_m$-averages are closer to the optical depth scale (see Fig. 1). In the deeper layers below the superadiabatic regime (SAR), the various averaging methods are practically indistinguishable. In the upper atmosphere the differences are smaller at higher $[\mathrm{Fe/H}]$ due to relatively small horizontal contrast, however, these increase significantly for lower metallicity. The averages $\langle 3D \rangle_z$ and $\langle 3D \rangle_m$ are marginally cooler than $\langle 3D \rangle_{\mathrm{Ross}}$ by $\sim 1 - 2\%$ at solar metallicity. In the metal-poor case $[\mathrm{Fe/H}] = -3.0$ the temperature stratifications are distinctively cooler, which will certainly influence the line formation calculations with $\langle 3D \rangle$ stratifications. Furthermore,
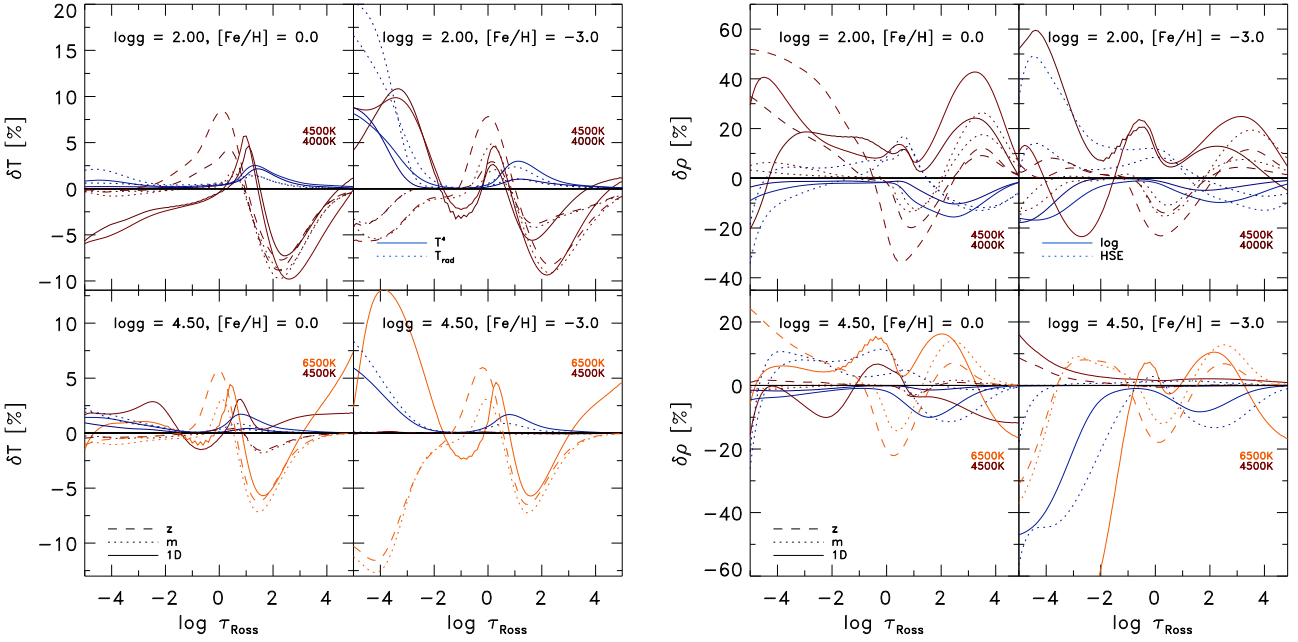
Fig. 1. Relative differences in the temperature (left) and density (right panel) stratification vs. the (averaged) Rosseland optical depth for various stellar parameters. The differences are relative to the Rosseland optical depth, i.e. $\langle 3D \rangle_{\bar{z}} - \langle 3D \rangle_{Ross}$. *Orange dashed lines*: averages on layers of constant geometrical height $\langle 3D \rangle_z$; *orange dotted lines*: averages on layers of constant column mass density $\langle 3D \rangle_m$; *orange solid lines*: 1D MLT models. In the left panel, the flux-weighted T^4-stratifications (*blue solid lines*) and the brightness temperatures T_{rad} (*blue dotted lines*) averaged on surfaces of constant Rosseland optical depth are also shown. The logarithmic density averages $\langle 3D \rangle_{Ross}^{log}$ (*blue solid lines*) and hydrostatic averages $\langle 3D \rangle_{Ross}^{HSE}$ (*blue dotted lines*) are shown in the right panel. Note the differences in the y-axes.

the differences increase with higher T_{eff} and lower $\log g$.

As mentioned earlier, in the atmospheres of late-type stars, minor temperature fluctuations are amplified disproportionally into large variations in continuum opacity κ_λ due to the strong T-sensitivity of the H$^-$-opacity ($\kappa_\lambda \propto T^{10}$, see Stein & Nordlund 1998). Therefore, surfaces of constant optical depth appear strongly corrugated in terms of the range of geometrical heights they span. The transformation to layers of constant optical depth will naturally even out these corrugated surfaces and, at the same time, smooth the temperature fluctuations, since the latter are the source of the former (see App. A.1). That temperature fluctuations become rather small on constant optical depth can be appreciated by looking at the temperature contrast and temperature distributions as a function of depth (see also Figs. 3 and 6). The superadiabatic regime (SAR) exhibits large-amplitude fluctuations as a result of the release of thermal and ionization energy at the photospheric transition, which are the reason for the observed enhanced differences between the averaging methods (see Sect. 4.1).

Steffen & Holweger (2002) found a beneficial mean $\langle T \rangle$-representation for the Sun in the flux-weighted temperature averages, T^4, taken on constant Rosseland optical depth from their 2D simulations. The idea behind this approach is that the T^4-averages render radiation-oriented T-stratifications, therefore resulting in 1D line profiles that are closer to the multi-dimensional ones (see also Steffen et al. 1995). To allow for a similar comparison for our models, we computed such average T^4-stratifications. In Fig. 1, the T_{Ross}^4-stratifications (blue solid lines) appear generally hotter at the top and in the SAR compared to the simple T-stratification. Averages

taken at the fourth power will weight higher values more, which leads to hotter average temperatures. This could lead to pronounced differences for molecular lines that form high up in the atmosphere. At solar metallicity the T^4-stratifications at the top are rather similar to the plain T-averages ($\sim 1-2\%$) in agreement with the findings of Steffen & Holweger (2002). This is different at lower metallicity ([Fe/H] = -3.0), namely the T^4-averages are clearly higher by $\sim 5-10\%$. At higher T_{eff} and lower $\log g$ the temperature differences are larger, in particular for the metal-poor giants, due to the enhanced temperature fluctuations (see Sect. 4.1).

Under the assumption of local thermodynamic equilibrium (LTE) and neglecting the effects of scattering, the source function is given by the Planck function, $S_\lambda = B_\lambda(T)$. Hence, within this approximation, we can consider the brightness temperature average T_{rad} defined earlier in Sect. 2.2 as a good representation of the mean temperature stratification from the point of view of the radiative emission properties: brighter parts in each depth layer are given larger weight with this averaging method. The differences between the average T_{rad} at 500 nm and average T-stratifications are displayed in Fig. 1 (blue dotted lines). Differences with stellar parameter are rather similar to the T^4-average, however, slightly more pronounced, in particular the metal-poor giants exhibit at the top hotter stratifications by up to $\sim 20\%$.

3.2. Density

In Fig. 1, we also illustrate the results of averaging in the case of the density stratifications. In the deeper interior, the different $\langle 3D \rangle$ models converge toward the same density stratification. In the SAR, below the optical surface at $\log \tau_{Ross} \gtrsim 0.0$,

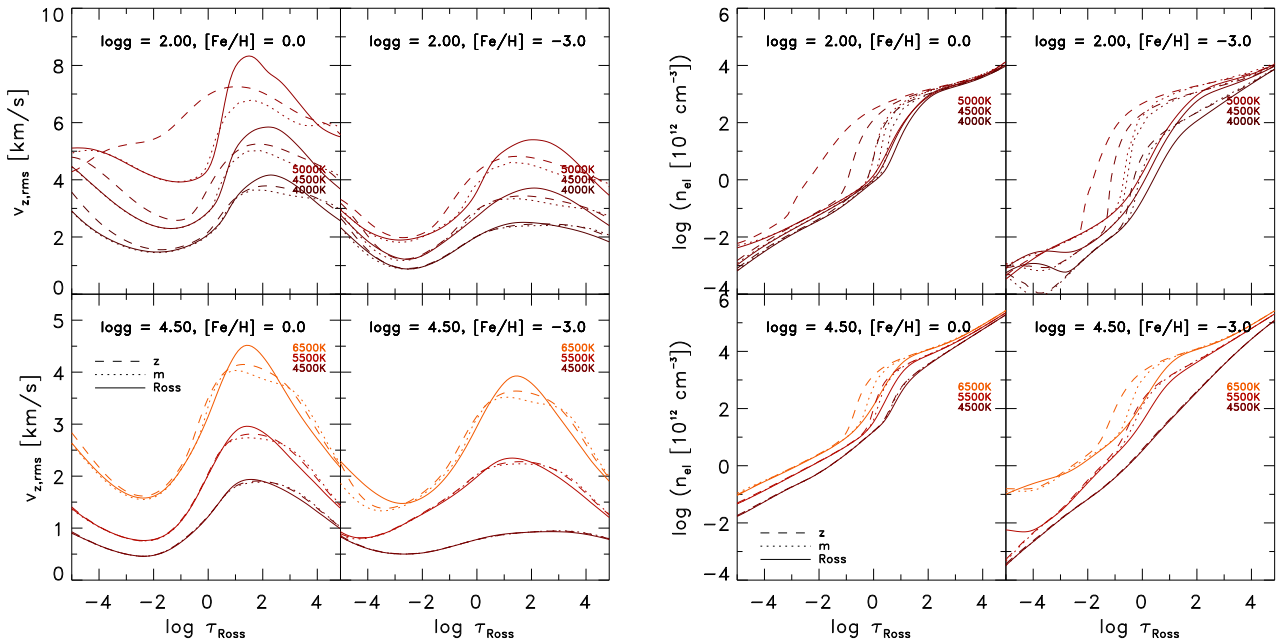**Fig. 2.** Root mean square (rms) of the vertical velocity $v_{z,\mathrm{rms}}$ (left) and mean electron number density n_{el} vs. optical depth (right panel). *Dashed lines:* $\langle 3D \rangle_z$ averages; *dotted lines:* $\langle 3D \rangle_m$; *solid lines:* $\langle 3D \rangle_{\mathrm{Ross}}$.

the geometrical averages $\langle 3D \rangle_z$ are smaller by up to $\sim 30\%$ compared to the $\langle 3D \rangle_{\mathrm{Ross}}$ averages, while at the top these are considerably denser by up to $\sim 40\%$. The differences increase towards higher T_{eff} and lower $\log g$. We find a different behavior in the metal-poor dwarfs, which turn lower towards the top after the initial increase ($\sim 10\%$). The density stratifications averaged on column mass density $\langle 3D \rangle_m$ are larger in the SAR, and in the upper layers closer to $\langle 3D \rangle_{\mathrm{Ross}}$. However, we find that at lower metallicity $\langle \rho \rangle_m$ they are smaller by up to $\sim 30\%$. We remark that thermal pressure shows qualitatively the same characteristics as the density.

The shape of the density distribution is symmetric and narrow on layers of constant column mass density, due to the exponential stratification of the atmosphere and to the additional damping of density fluctuations on the column mass scale (see Fig. 6). Hence, the $\langle 3D \rangle_m$ averages feature the smallest contrast and density ranges, which, on the contrary, are usually largest for geometrical averages $\langle 3D \rangle_z$; for the $\langle 3D \rangle_{\mathrm{Ross}}$ averages, these are noticeably reduced due to the mapping onto the optical reference depth scale (Fig. 3). Overall, the density fluctuations at the top of the $\langle 3D \rangle_{\mathrm{Ross}}$ stratifications are similarly as small as those by $\langle 3D \rangle_m$ and of the order of $\sim 20\%$; however, for metal-poor dwarfs they reach up to $\sim 80\%$ (see Fig. 3). As shown in Sect. 4.3, we find that the corrugation of the layers of constant optical depth in the upper part of 3D model stellar atmospheres at lower metallicity increases considerably towards higher T_{eff} due to an enhanced T-contrast by the so-called reversed granulation (see Rutten et al. 2004). This in turn broadens the density distribution during the remapping to the optical depth scale, shifting the mean density value and leading to the observed deviations between $\langle \rho \rangle_{\mathrm{Ross}}$ and $\langle \rho \rangle_m$ at lower metallicity (see App. A.1), which will affect the $\langle 3D \rangle$ line formation calculations.

The highly stratified structure of stellar atmospheres features an exponential decrease with height. Linear density averages

will therefore tend to give more weight to larger density values, leading to a systematic overestimation of the mean densities. For this reason we consider the logarithmic averages $\langle \rho \rangle_{\mathrm{log}}$, which we compare to the linear ones in Fig. 1 (blue solid lines). As expected, we find the logarithmic ρ-averages being smaller than the linear ones, with the difference between the two increasing with higher T_{eff} and lower $\log g$ by up to $\sim 30\%$. The mean densities in the upper layers are smaller by $\sim 10\%$ and $\sim 40\%$ at solar and low metallicity, respectively. For quantities that vary more moderately (e.g. temperature) the differences between logarithmic and linear averaging are rather small.

The transformation to constant optical depth and the subsequent averaging will change the physical self-consistency as shown in App. A.2. To rectify this, we followed the recommendation by Uitenbroek & Criscuoli (2011) and computed also ρ-stratifications, which are enforced to be in hydrostatic equilibrium, $\langle \rho \rangle_{\mathrm{HSE}}$ (blue dotted lines in Fig. 1). These deviate rather significant from the plain $\langle \rho \rangle$-stratifications, in particular at the top. Incidentally, we note however that due to their dynamic nature and the effects of convective flows and turbulent pressure the 3D models themselves are not strictly speaking in hydrostatic equilibrium at any one time.

In Fig. 1 (both panels), we compare also the differences between $\langle 3D \rangle_{\mathrm{Ross}}$ and the 1D MLT models. These exhibits qualitatively similar difference as the geometrical averages, but in the metal-poor case the 1D models are distinctively hotter due to the enforcement of radiative equilibrium.

3.3. Electron number density

We find large differences among the various averages of the electron number density, n_{el}, which we show in Fig. 2 (right panel). In the SAR the geometrical averages $\langle n_{\mathrm{el}} \rangle_z$ are distinctively larger than the optical averages $\langle n_{\mathrm{el}} \rangle_{\mathrm{Ross}}$, while the column mass density averages $\langle n_{\mathrm{el}} \rangle_m$ are found in between the two. The deviations increase for higher T_{eff} and lower $\log g$ considerably,

while at lower T_{eff} the differences are significantly smaller. We show in App. A.1 that the interpolation to a new reference depth scale changes the statistical properties by redistributing properties from different heights, hence the resulting mean horizontal average will look different depending on the reference depth scale. This effect seems to be most pronounced in the case of electron density.

To determine the ionization fraction in spectral line calculations, the electron number density is either already provided by the model atmosphere, or looked up from an EOS using the independent thermodynamic variables (typically (T, p), or (T, ρ)). The latter has to be done carefully in the case of the $\langle 3D \rangle$ models, since, besides potential differences in the EOS compared to the one used for calculating the model atmosphere, electron densities derived from the EOS based on averaged independent variables, $n_{el}^{EOS} = n_{el}(\langle T \rangle, \langle p \rangle)$, can deviate significantly from the more physically consistent averaged $\langle n_{el} \rangle$ (see App. A.3).

3.4. Vertical velocity

It is worthwhile to compare how the vertical velocity, $v_{z,rms}$, changes with the respective averaging methods. For comparison, we show in Fig. 2 (left panel) the rms of the vertical velocity. In the upper layers, we find the $v_{z,rms}$ on geometrical averages to be higher compared to other averages, while it is lower in the deeper layers. On optical depth the peak in $v_{z,rms}$ below the surface is rather symmetric and slightly higher, while for averages on geometrical height and column mass density their peaks are flatter and more skewed towards higher layers, and the peak location is realized in slightly upper layers. For lower T_{eff} and higher $\log g$ the differences diminish more and more, so that for the coolest models, the difference are small. The difference in the velocity arise as well due to the redistribution of velocity during the mapping to the new reference depth scale (see App. A.1).

4. Statistical properties

In order to explore the origins of the differences among the various average $\langle 3D \rangle$ structures and the resulting ramifications for line formation calculations, we discuss here the statistical properties of the temperature, density, and velocity stratifications. Since the statistical properties of $\langle 3D \rangle_{500}$ and $\langle 3D \rangle_{Ross}$ are rather similar, we will focus only the latter.

4.1. Contrast

The 3D RHD models exhibit usually a broad range of values at a given height due to the fluctuations arising from the convective motions. The amplitude of these fluctuations can be quantified using the root-mean-square of the relative deviation from the mean,

$$\delta X_{rms} = \sqrt{\Sigma_{i=1}^{N} \left(X_i - \bar{X} \right)^2 / \left(N \bar{X}^2 \right)}, \tag{3}$$

which we refer to as the *contrast* ($\bar{X}$ is the mean value of X). It is equal to the normalized standard deviation, i.e. $\delta X_{rms} = \sigma_X / \bar{X}$.

The translation to another reference depth scale changes the statistical properties as variables are remapped, which in turn is reflected in changes in contrast. Among the various averaging methods, geometric averages $\langle 3D \rangle_z$ typically feature the highest contrast. We also find that the level of fluctuations generally increases with increasing T_{eff} and decreasing $\log g$.
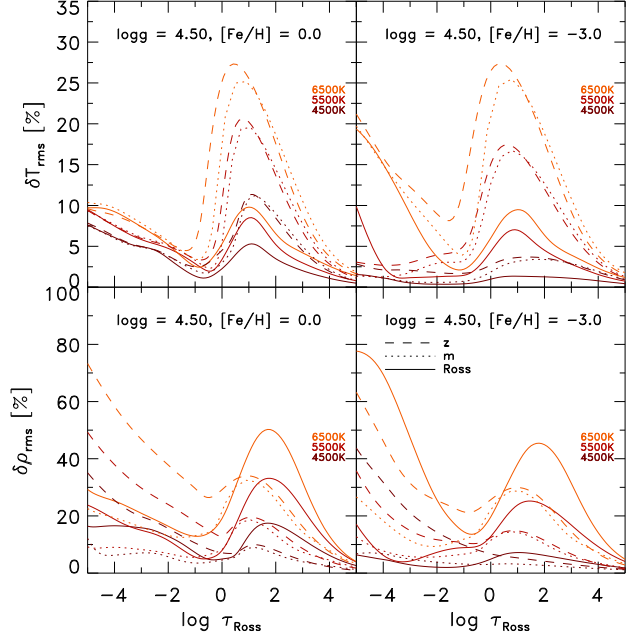

Fig. 3. Temperature (top) and density (bottom) contrasts vs. averaged Rosseland optical depth. *Dashed lines*: $\langle 3D \rangle_z$ averages; *dotted lines*: $\langle 3D \rangle_m$; *solid lines*: $\langle 3D \rangle_{Ross}$.
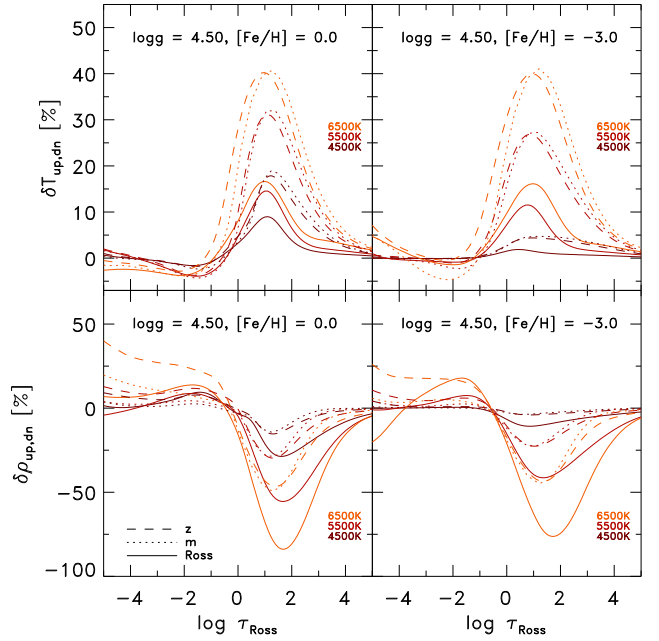

Fig. 4. Similar as Fig. 3 but showing the relative difference between averages in up and downflows, $\delta T_{up,dn}$ and $\delta \rho_{up,dn}$.

The largest contrast is typically prevailing in simulations with the highest T_{eff} in the SAR, ∇_{sad}^{peak}, where also the velocity peak, $v_{z,rms}^{peak}$, is located. These arise due to the photospheric transition from convective to radiative energy transport, and the resulting overturning of the entropy-depleted plasma. At the top of the convection zone, the fluctuations reach a minimum; they also decrease towards the bottom of the model atmosphere.

In Fig. 3, top and bottom panel, we show the temperature and density contrasts, $\delta T_{\rm rms}$ and $\delta\rho_{\rm rms}$, respectively. In the case of the optical depth $\langle 3D\rangle_{\rm Ross}$ the temperature contrast is significantly reduced compared to the other reference depth scales ($\delta T_{\rm rms}^{\rm peak}$ reduced by a factor of ~ 3), while the density contrast is slightly enhanced ($\delta\rho_{\rm rms}^{\rm peak} \sim 20 - 60\,\%$ compared to $10 - 50\,\%$). For averages on column mass density $\langle 3D\rangle_m$, $\delta\rho_{\rm rms}$ is lower, in particular in the upper layers, and $\delta T_{\rm rms}$ is slightly smaller compared to the $\langle 3D\rangle_z$ case. Fluctuations of variables which correlate with the new reference depth scale will be reduced during the transformation. As the translation to layers of constant optical depth partly evens out the corrugated τ-isosurface, fluctuations of the opacity κ_λ will be reduced, since the dominant H^- opacity is very sensitive to temperature. Therefore, the temperature fluctuations are also smoothed out. Layers of constant column mass density will similarly suppress density variations (see App. A.1). At the top, $\delta\rho_{\rm rms}$ is rather similar between $\langle 3D\rangle_m$ and $\langle 3D\rangle_{\rm Ross}$ in the case of the solar metallicity ($\delta\rho_{\rm rms}^{\rm top} \sim 40\,\%$), however, at lower metallicity, [Fe/H] = −3.0, we find a considerable disparity with $\delta\rho_{\rm rms}^{\rm top} \sim 80\,\%$.

The thermal stratification in the upper atmosphere is determined by adiabatic cooling due to mechanical expansion and radiative heating due to spectral line re-absorption (Asplund et al. 1999; Collet et al. 2007). In metal-poor stars, radiative re-heating in upper layers is significantly reduced due to the weakness of spectral line features, while the mechanical expansion cooling term is virtually unaffected. The reversed granulation takes place at increasingly lower geometrical height with higher $T_{\rm eff}$ and lower $\log g$, causing the distribution of the thermodynamic variables to become increasingly broader and more skewed (see 4.3). This is the reason for the enhancement in $\delta T_{\rm rms}$ and $\delta\rho_{\rm rms}$ towards the top boundary in metal-poor simulations in Fig. 3. Replicating the results of full 3D line formation calculations in low metallicity stars with $\langle 3D\rangle$ models is therefore challenging, since the averages have to account correctly for such temperature and density fluctuations. Interestingly, the temperature contrast saturates at 6500 K, similar to the saturation of the intensity contrast shown in our previous work (see Fig. 10 in Paper I).

The strength of spectral lines depends sensitively on temperature, and the remapping to constant optical depth decreases $\delta T_{\rm rms}$, making $\langle T\rangle$ closer to $\langle T\rangle_{\rm rad}$. However, the transformation to layers of constant optical depth exhibits the side effect of redistributing the other variables, too, in particular the gas density; $\delta\rho_{\rm rms}$ is thus much larger compared to averages on column mass density, due to the additional influence of opacity on the depth scale (see Sect. 2.2). This in turn will likely affect the line formation calculations with the different $\langle 3D\rangle$ models.

The strong contrast in the upper part of the convection zone ($\log \tau_{\rm Ross} \geq 0$) is induced by the large amplitude fluctuations due to the radiative energy losses at the photosphere and the asymmetry of the up- and downflows, which we discuss further in Sect. 4.2. An interesting aspect is that the contrast in thermodynamic variables is rather similar to the rms of the vertical velocity (Fig. 2), which is indicative of the correlation between the mass flux and the fluctuations in the thermodynamic variables. Namely, vertical velocity is generated by density contrast $\delta\rho$ via to the buoyancy force, $f_B = -g\delta\rho$, which results from an imbalance of pressure and gravity terms in the hydrodynamical equation for conservation of momentum (see Paper I) in the highly stratified atmosphere. Lighter fluid
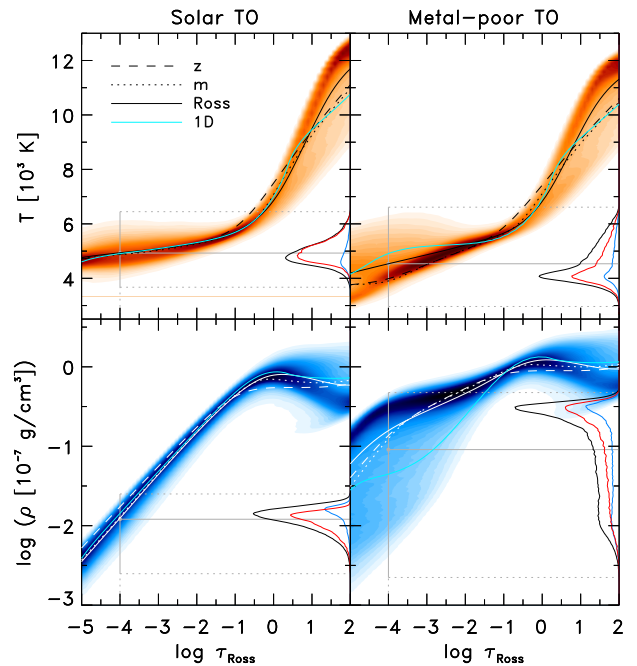
Fig. 5. Histogram of the temperature (top) and density (bottom) vs. optical depth for the TO simulation ($T_{\rm eff} = 6500\,{\rm K}/\log g = 4.0$) with solar and sub-solar metallicity ([Fe/H] = −3.0). Additionally, the histogram of a single layer ($\log \tau_{\rm Ross} = -4.0$) is indicated for the whole layer (black) and separated in up- and downflows (blue and red, respectively). *Dashed lines*: $\langle 3D\rangle_z$ averages; *dotted lines*: $\langle 3D\rangle_m$; *solid lines*: $\langle 3D\rangle_{\rm Ross}$; *blue solid lines*: 1D MLT models.

elements ($\delta\rho < 0$) experience positive buoyancy and thus upward acceleration, while denser elements ($\delta\rho > 0$) experience negative buoyancy and are pulled downward. Buoyancy forces will vanish eventually, when the density of the up- or downflowing element levels with the surrounding gas.

The entropy contrast $\delta s_{\rm rms}$ (not shown here), depicts qualitatively a very similar dependence with stellar parameter and reference depth scale as $\delta T_{\rm rms}$. Both are rather similar on optical depth, while for the averages $\langle 3D\rangle_z$ and $\langle 3D\rangle_m$ the overall amplitude is a factor ~ 2 smaller. In Paper I, we showed that the convective energy flux depends on the entropy jump, density and vertical velocity. Interestingly, here we also find additional *scaling relations* concerning the peak contrast in entropy, $\delta s_{\rm rms}^{\rm peak}$, and density, $\delta\rho_{\rm rms}^{\rm peak}$, with the vertical peak velocity $v_{z,{\rm rms}}^{\rm peak}$. This can be interpreted as convective driving, where the radiative losses generate large fluctuations in the entropy, temperature and density.

For the different averaging methods the variations of the minimum-maximum range for the temperature and density are qualitatively very similar to the contrast (even though with larger amplitudes $\sim 5 - 8$), therefore, we refrain from discussing these explicitly.

4.2. Upflows and downflows

The properties of the convective motions in stellar atmospheres are highly asymmetric in up- and downflows. The upflows overshoot into the photosphere leading to non-thermal Doppler shifts imprinted on spectral line features. We first compute the mean values of various variables separately for up and downflows

based on the sign of the velocity at a given height; we then determine the relative difference between up- and downflows with $\delta X_{\mathrm{up,dn}} = (X_{\mathrm{up}} - X_{\mathrm{dn}})/\bar{X}$ (Fig. 4). As expected, the buoyant upflows are hotter and lighter compared to the subsiding downflows. Furthermore, the asymmetries are especially pronounced in the convection zone below the optical surface. Above the photosphere, the convective motions decay quickly and the asymmetries in $\delta T_{\mathrm{up,dn}}$ and $\delta \rho_{\mathrm{up,dn}}$ are distinctively smaller. The remaining asymmetries at the top stem from reverse granulation.

The convective flows in granules, slow and almost laminar, radiate away their energy and overturn into the intergranular lanes characterized by cool, dense, narrow turbulent downdrafts. Most of the fluctuations are caused by these turbulent downdrafts blending with granules, generating vortex tubes around the latter. It is remarkable that, across all stellar parameters, the filling factor of the up- and downflow in the convection zone remains rather constant, with $f_{\mathrm{up}} \sim 2/3$ and $f_{\mathrm{dn}} \sim 1/3$, respectively (see Paper I).

$\delta T_{\mathrm{up,dn}}$ is reduced and $\delta \rho_{\mathrm{up,dn}}$ is enhanced on the optical reference depth scale $\langle 3D \rangle_{\mathrm{Ross}}$ compared to the other averages. The column mass density show a smaller asymmetry in density. This behavior, similar to what we discussed earlier for the temperature and density contrasts, is not entirely surprising, since the fluctuations are caused by the presence of the up- and downflows (see also A.1).

4.3. Histograms

In Fig. 5, we illustrate temporally averaged histograms of the temperature, $p(T)$, and density distributions, $p(\rho)$ for the TO simulation with two different [Fe/H], in order to illustrate the differences in the statistical properties. The histogram of the metal-poor case differs substantially in upper layers from the solar one. Furthermore, in Fig. 6, we show $p(T)$ and $p(\rho)$ in the upper layers ($\langle \log \tau_{\mathrm{Ross}} \rangle = -4.0$) for dwarf models with different T_{eff} and [Fe/H]. In both cases we compare the distributions on constant geometrical height z, constant column mass density m and constant Rosseland optical depth τ_{Ross}.

At solar metallicity (dashed lines), the temperature distributions are rather narrow and symmetric. With increasing T_{eff} the average T is as expected higher and the width of the distribution broadens slightly. The mean values are rather similar between the different $\langle 3D \rangle$ methods and in principle indistinguishable, which is also in agreement with Fig. 1. Furthermore, the mean values are located very close to the mode.

At [Fe/H] = −3.0 (solid lines), the temperature distributions change considerably. While at cooler T_{eff} the shape is rather narrow and symmetric, for $T_{\mathrm{eff}} \geq 5500\,\mathrm{K}$ we find a distinct broadening of the T-distribution on geometrical reference depth scale $\langle 3D \rangle_z$ that is given by a long tail at high T and a decreasing peak at lower T (see Figs. 5 and 6). In the column mass density averages $\langle 3D \rangle_m$ the temperature peak is slightly more pronounced at higher T_{eff}, while the high-T tail is slightly reduced. The situation is rather different for the averages on Rosseland optical depth $\langle 3D \rangle_{\mathrm{Ross}}$, where we find that the temperature peak drops faster towards higher T_{eff}, and at 7000 K the T-distribution show a rather unimodal distribution. The mean values disagree at higher T_{eff} between the different reference depth scales.

The density distributions behave rather differently depending on the reference depth scale. On $\langle 3D \rangle_z$ the histograms are in general slightly skewed with a fat tail towards lower ρ for all metallicities (Figs. 5 and 6). The density distributions for

the averages on column mass density are quite symmetric and narrow for both solar and low metallicity. At solar metallicity, the density histogram on constant optical depth are narrower and higher than the geometrical analogs, but skewed in contrast to $\langle 3D \rangle_m$. In the metal-poor case, $\langle p(\rho) \rangle_{\mathrm{Ross}}$ becomes very narrow and symmetric at lower T_{eff}, but towards higher T_{eff} we find the ρ-distribution to be also broader. The mean density stratification varies considerably among the different averaging methods.

As mentioned above, adiabatic cooling due to mechanical expansion and radiative re-heating are competing with each other in the upper photosphere and contributing to the phenomenon of reversed granulation. At lower metallicity, the reversed granulation is enhanced thereby the optical depth is increasingly strongly corrugated towards higher T_{eff}, which in turn will amplify the differences during the translation to the optical depth scale. This leads to the systematical broadening in the statistical distribution that we encounter at lower metallicity.

5. Spectral line formation: $\langle 3D \rangle$ and 3D LTE calculations

In order to explore the differences between the line formation based on $\langle 3D \rangle$ and full 3D models, we have chosen a set of representative models consisting of a main-sequence (MS) star ($T_{\mathrm{eff}}/\log g = 5777\,\mathrm{K}/4.44$), a turn-off (TO) star (6500/4.0), a red-giant (RG) star (4500/2.0) and a dwarf (4500/5.0). For all these models, we considered besides the solar metallicity also metal-poor analogs with [Fe/H] = −3.0.

5.1. 3D line formation calculations

We used the 3D radiative transfer code Scate (Hayek et al. 2011) to calculate full 3D synthetic spectral line disk-center intensity and flux profiles with 3D Stagger model atmospheres. Scate assumes local thermodynamic equilibrium (LTE); furthermore, in the present work, we also neglected the effects of scattering, i.e. we approximated the source function with the Planck function, $S_\lambda = B_\lambda$. We caution that LTE is in general a poor approximation, especially for Fe I spectral line formation calculations at low [Fe/H] (e.g. Bergemann et al. 2012), which should be kept in mind for analyzing the LTE-based abundance corrections presented here.

For the sake of consistency, we used the same EOS (Mihalas et al. 1988) and continuum opacity data (from the MARCS package; see Gustafsson et al. 2008) as in the 3D Stagger simulations.

To reduce the computational costs for line formation calculations, we consider a subset of $N_t = 20$ temporally equidistant snapshots –the same used for the temporal $\langle 3D \rangle$ averages– sampling the entire time-spans of the individual 3D simulation sequences. Additionally, we reduce the horizontal spatial resolution from $N_x N_y = 240^2$ to 60^2 by considering only every 4th column in each horizontal direction. Test calculations carried out at full resolution show that differences are negligible for all practical purposes (see Asplund et al. 2000a). Concerning the vertical direction, while we did not subsample the number of depth points, we considered only the layers with $\min(\log \tau_{\mathrm{Ross}}) \leq 3.0$. The resulting disk-center intensity and flux profiles are spatially and temporally averaged, and then
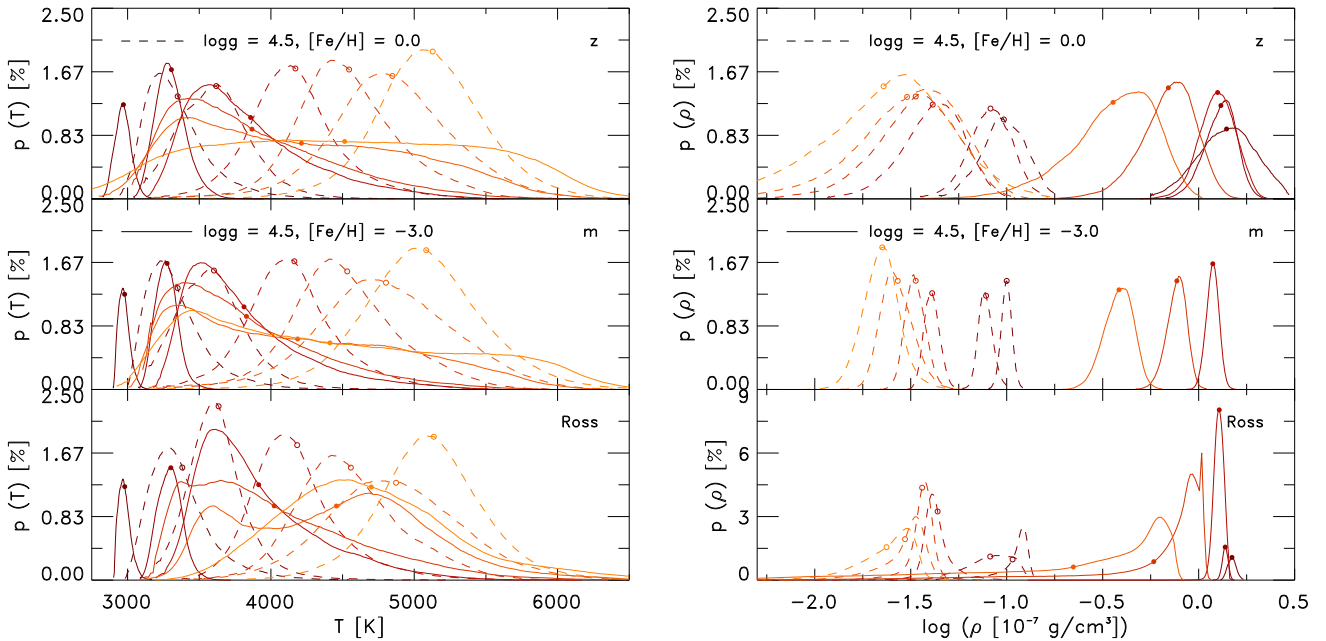
Fig. 6. Histograms of the temperature (left) and density (right panel) distributions taken at $\langle \log \tau_{\mathrm{Ross}} \rangle = -4.0$. We show the histograms averaged on constant geometrical height (top), column mass density (middle) and Rosseland optical depth (bottom). The surface gravity of displayed models is $\log g = 4.5$ and the metallicity is solar (dashed lines) and sub-solar with [Fe/H] = -3.0 (solid lines). The mean values are indicated by open and filled circles for [Fe/H] = -3.0 and 0.0, respectively.

normalized with the respective continuum intensity or flux.

To illustrate systematically the differences between $\langle 3D \rangle$ and 3D line formation, we computed fictitious atomic lines for neutral and singly ionized iron, Fe I and Fe II, for the selected STAGGER-grid models and metallicities. All lines are defined at the same wavelength, $\lambda = 500$ nm, and we consider two lower-level excitation potentials, $\chi_{\mathrm{exc}} = 1.0$ and 4.0 eV. Furthermore, we vary the oscillator strength, $\log gf$, in order to cover a range of line strengths, from weak to partly saturated lines, with equivalent widths from $W_\lambda = 5$ to 80 mÅ. We assume an iron abundance of $\log \epsilon_{\mathrm{Fe}} = 7.51$ (Asplund et al. 2009) and $\log \epsilon_{\mathrm{Fe}} = 4.51$, for the solar-metallicity and [Fe/H] = -3.0 case, respectively.

The spectral line calculations with $\langle 3D \rangle$ models are also performed with SCATE, to guarantee a consistent comparison. SCATE employs atmospheric structures on geometrical height and computes the optical depth, τ_λ, for the individual line. Therefore, we provide the geometrical height by integrating $dz = d\langle\tau_\lambda\rangle / \langle\kappa_\lambda\rangle$, which is of course unnecessary for $\langle 3D \rangle_z$. Furthermore, tests revealed that including just an averaged velocity, e.g. $|\mathbf{v}|/3$, is insufficient to reproduce the influence of the 3D velocity field on the line shape. Analyzing the influence of the velocity field on the line formation surpasses the scope of the present work; therefore, we will explore this aspect in a separate study. In this paper, for the calculations with $\langle 3D \rangle$ models we neglected the information about the actual velocity field, and assumed instead a fixed microturbulence of $\xi_{\mathrm{turb}} = 1.0$ km/s for all considered stellar parameters.

Since the line formation calculations with $\langle 3D \rangle$ models are obviously much faster, we use the $\langle 3D \rangle_{\mathrm{Ross}}$ averages first to estimate the $\log gf$ range, which would result in the designated range in W_λ. We then consider ten equidistant $\log gf$ values within that range for the $\langle 3D \rangle$ and full 3D models. Finally, we interpolate the curves of growth ($\log gf$ vs. W_λ) using

a spline interpolation, and retrieve the $\Delta \log gf$ difference between $\langle 3D \rangle$ and 3D synthetic lines at a given equivalent width, i.e. $\Delta \log gf = \langle 3D \rangle - 3D$. For trace elements, changes in line strength due to $\Delta \log gf$ are equivalent to changes due to abundance variations $\Delta \log \epsilon$; hence, the $\Delta \log gf$ differences can be interpreted as $\langle 3D \rangle - 3D$ abundance corrections. With four fictitious lines and four representative models with two metallicities, we covered 32 cases in total.

Full 3D line profiles are marked by line shifts and asymmetries due to the non-thermal Doppler broadening introduced by the up- and downflows of the convective motions, which are present in the photosphere due to overshooting (Asplund et al. 2000a). In 3D RHD modeling, the velocity field emerges naturally from first principles; the buoyant hot rising plasma in the granules blue-shifts the line, while the fast downdrafts introduce a red-shift. Besides the convective motions, another source of line broadening are the inhomogeneities in the thermodynamic independent variables, ρ and T. The ascending granules are hotter and less dense, compared to the downdrafts (see Fig. 4). The velocities and inhomogeneities prevailing at formation height of the individual lines will lead to line shifts and asymmetries. The $\langle 3D \rangle$-based lines are symmetric without any shifts, however, we can compare the equivalent widths of lines from calculations based on full 3D models and on the different average stratifications.

We probe different formation heights with the parameters of our fictitious lines. The Fe II lines form deeper in the atmosphere, closer to the continuum forming layers, while the Fe I lines are more sensitive to the intermediate heights of the atmosphere. Spectral lines with lower (higher) excitation potential form at smaller (larger) optical depths. We showed in Sect. 3 that the metal-poor model stellar atmospheres exhibit rather different temperature stratification at the top depending on the av-
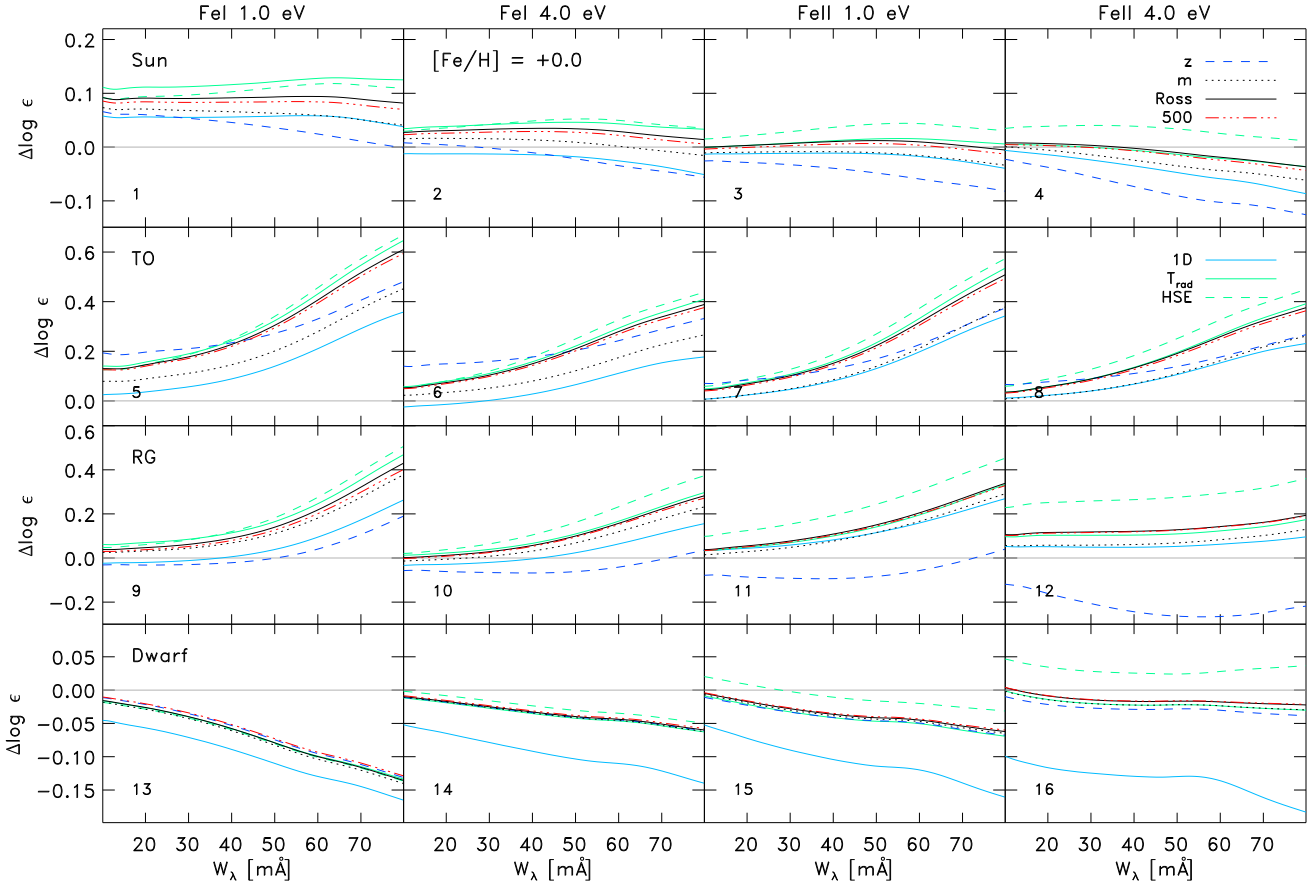
Fig. 7. Overview of the $\langle 3D \rangle - 3D$ line formation differences given in abundances displacement $\Delta \log \varepsilon$ vs. equivalent width W_λ for the Fe I and Fe II fictitious spectral lines with the excitation potentials $\chi_{\mathrm{exc}} = 1.0$ and 4.0 eV. Showing the representative selection including the Sun, TO, RG and dwarf (from top to bottom; notice their different ordinates). The averages on layers of constant geometric height $\langle 3D \rangle_z$ (blue dashed), constant column mass density $\langle 3D \rangle_m$ (black dotted), constant Rosseland optical depth $\langle 3D \rangle_{\mathrm{Ross}}$ (black solid lines) and at 500 nm $\langle 3D \rangle_{500}$ (red dashed triple-dotted) are indicated. For these averages, we show the solar (black) and sub-solar (blue lines) metallicity. Furthermore, we show 1D models (blue solid), $T_{\mathrm{rad}}^{\mathrm{Ross}}$-averages (green solid) and $\langle 3D \rangle_{\mathrm{Ross}}^{\mathrm{HSE}}$ (green dashed lines). The microturbulence of $\xi_{\mathrm{turb}} = 1.0$ km/s has been used throughout.

eraging method, consequently the latter should show the largest differences between the $\langle 3D \rangle$ models.

5.2. Comparison of $\langle 3D \rangle$ and 3D line formation

We show an overview of the differences between the $\langle 3D \rangle$ and the full 3D calculations in Figs. 7 and 8. The first noticeable observations are the systematic trends in form of a slope towards higher line strength, which are due to the fixed value of the microturbulence, ξ_{turb}, with 1 km/s in the $\langle 3D \rangle$ models. An increasing slope with line strength indicates an underestimation of ξ_{turb} in particular for the TO and RG (see panel 5 to 12 in Fig. 7 and 21 to 28 in Fig. 8). By contrast, in cool dwarfs, the adopted ξ_{turb} seem to be overestimated. These findings are in agreement with comparisons of 1D models with observations (e.g. Edvardsson et al. 1993; Bensby et al. 2009). We tested this by applying a number of ξ_{turb} values[2], which showed that a fine-tuning can rectify the present slope. However, for the sake of clarity, we prefer to limit the already large number of stellar and line parameters to just a single ξ_{turb} (the calibration of the

microturbulence will be the subject of a separate study).

Weak lines are insensitive to ξ_{turb} yet they show variations in strength, which can be attributed to differences in the mean $\langle 3D \rangle$ stratifications of temperature and density. Interestingly, when one compares this regime between the different averages in Fig. 7, the averages on column mass density (black dotted lines) are often the closest to the full 3D spectral lines and perform in this respect often better than the averages on constant optical depth (blue dashed lines). The stratification on constant optical depth at 500 nm (red dashed triple-dotted lines) shows always spectral line features slightly closer to the full 3D case compared to the Rosseland optical depth (black solid lines). However, this is due to the fact that we chose our fictitious iron lines at 500 nm, which leads to an inherent advantage of $\langle 3D \rangle_{500}$ over $\langle 3D \rangle_{\mathrm{Ross}}$. The geometrical averages show large deviations in the case of the TO and RG star at solar metallicity (see panels 5 to 12).

The differences in the metal-poor case (Fig. 8) are clearly larger than the solar metallicity models (Fig. 7). It is obvious that $\langle 3D \rangle$ models at low [Fe/H] struggle to reproduce the 3D case properly, in particular Fe I lines with small excitation potential; and the differences are particularly pronounced for the hotter metal-poor TO stars (panel 21). This is in accordance with our findings from Sect. 3 and 4: at low metallicity and high

---

[2] We find a reduction of the slope in the curve-of-growth with $\xi_{\mathrm{turb}} = 0.5, 1.5, 2.0$ km/s for the dwarf, RG and TO models respectively (while a fine-tuning could flatten it completely).
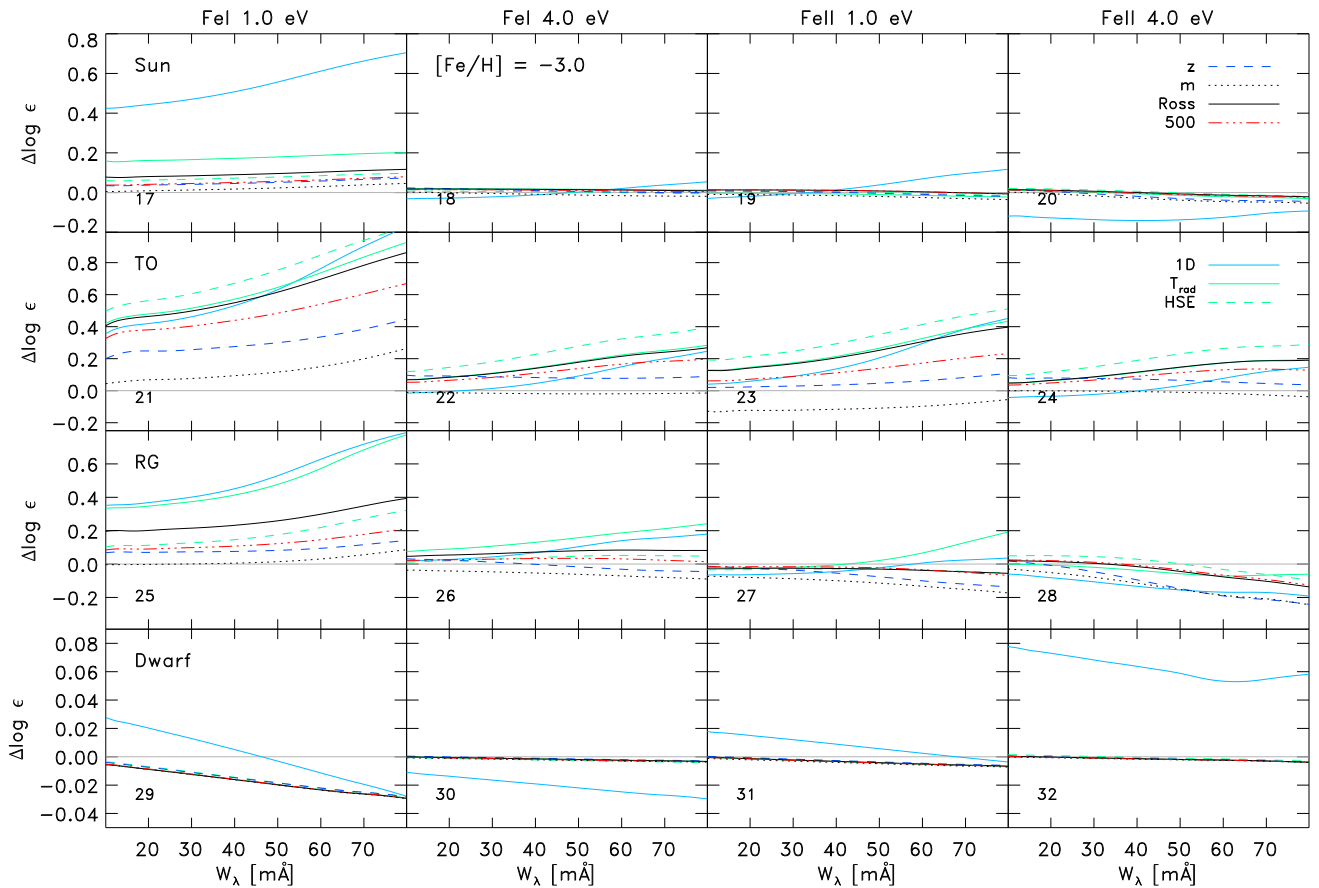
Fig. 8. Similar as Fig. 7 but showing overview of the abundance corrections for metal-poor models. Note the larger ranges of the *y*-scales.

T_{eff} the differences in the statistical properties among the various $\langle 3D \rangle$ averages increases at low [Fe/H]. In particular the widths of the temperature and density distributions becomes broader at lower metallicity (Fig. 6), and their mean values become increasingly less well-defined in its statistical representation. The reason for the broadening is the enhanced contrast of the reversed granulation due to the reduced radiative re-heating with weak spectral line features at low metallicity (see App.A.1).

In order to facilitate an overall comparison between the different averages with respect to line formation, we show in Fig. 9 (left) the mean abundance deviations for weak lines that are determined between $W_\lambda = 5 - 20$ mÅ. For the model representing the Sun, the differences between $\langle 3D \rangle$ and 3D are in general small: $\lesssim 0.1$ dex. For the TO stars at solar [Fe/H] the differences are considerably larger: $\lesssim 0.2$ dex. We find the largest deviations for Fe I lines with small excitation potential $\chi_{\text{exc}} = 1.0$ eV, which are the most temperature sensitive; in particular the geometrical averages exhibit strong differences. At lower metallicity, the differences increase in particular for the TO and RG model with $\lesssim 0.4$ dex, and the $\langle 3D \rangle$ on optical depth shows the largest deviation for metal-poor TO star. In general the deviations become smaller at higher χ_{exc} and for Fe II lines. The dwarfs show very little differences compared to the full 3D case. These models exhibit the smallest velocities and temperature contrast with the mean stratifications resembling closely the 1D models based on same EOS and opacities.

The averages on column mass density $\langle 3D \rangle_m$ exhibit typically the best performance, in particular at low metallicity.

The geometrical averages $\langle 3D \rangle_z$ exhibit large deviations (in agreement with Uitenbroek & Criscuoli 2011), especially for the TO stars. When one considers the comparison of the temperature and density in Fig. 1, then one can deduce that the models with cooler stratifications are closer to the full 3D line strength. The averages on constant optical depth perform surprisingly poorly, in particular for low excitation Fe I for the metal-poor TO star.

The resulting spectral line features with the logarithmic averages $\langle 3D \rangle_{\log}$ are similar to plain $\langle 3D \rangle_{\text{Ross}}$ (therefore we refrain from showing the latter), while averages enforcing hydrostatic equilibrium, $\langle 3D \rangle_{\text{HSE}}$, are clearly failing to reproduce closely the results from 3D line formation (similar to Uitenbroek & Criscuoli 2011), and leading to rather large errors in the line formation, in particular for the metal-poor TO model (Fig. 8). Furthermore, both the flux-weighted and brightness-temperature averages, T^4 and T_{rad}, are in general very close to the plain average, but often slightly less accurate, which is a somewhat surprising result (see T_{rad} in Fig. 9).

Another meaningful way to test the performance of the different averages can be accomplished by comparing the deviation of the center-to-limb variation (CLV) of the continuum intensity. In Fig. 9, we show the differences of the continuum intensity, $\delta I_\mu = (I_\mu^{\langle 3D \rangle} - I_\mu^{3D})/I_\mu^{3D}$, i.e. between the $\langle 3D \rangle$ and full 3D models. We find in general that the $\langle 3D \rangle$ models overestimate the continuum intensity at disk center ($\mu = 1$), while towards the limb ($\mu = 0.2$) the $\langle 3D \rangle$ underestimate often the intensity. The deviations of the different averages are similar to
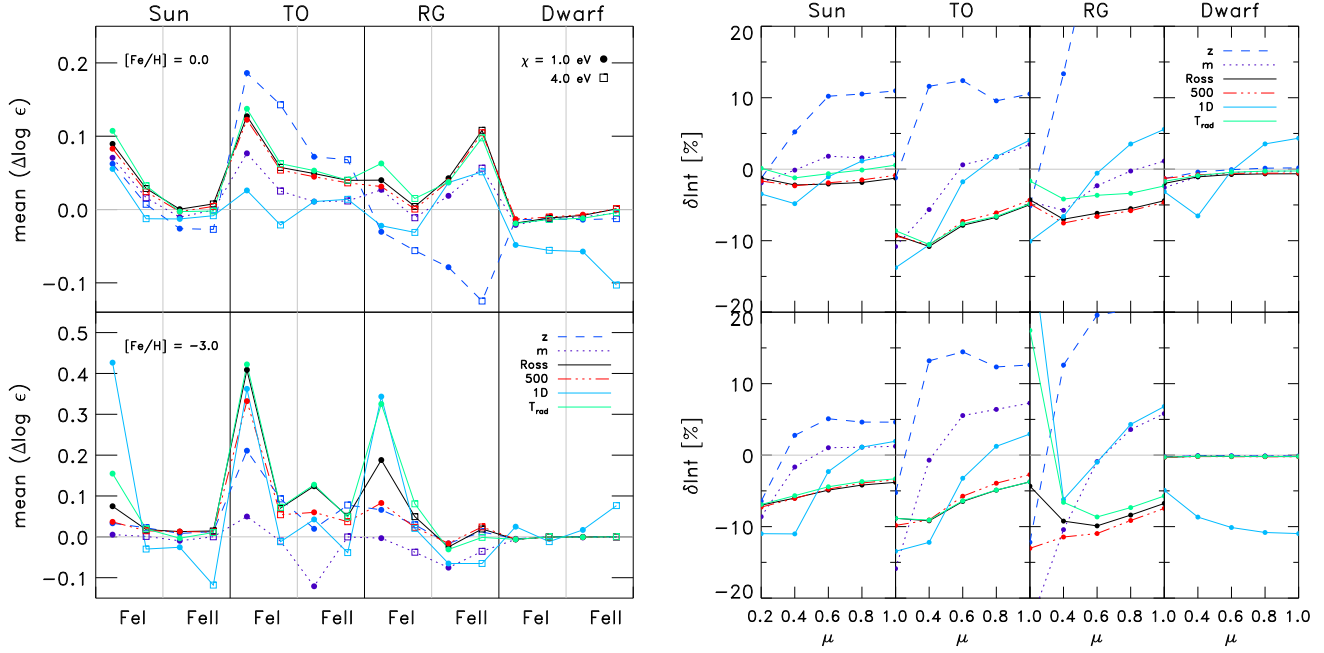
Fig. 9. In the left Figure the mean $\Delta \log \varepsilon$ (evaluated between $5 - 20\,\text{m\AA}$) is illustrated against Fe I and Fe II given at $\chi_{\text{exc}} = 1.0$ and $4.0\,\text{eV}$ for the different selected models. In the right Figure, the relative difference with $\langle 3D \rangle - 3D$ of the continuum intensity, δI_μ, vs. μ angle is displayed. Both Figures include the solar metallicity (top) and the metal-poor (bottom) case, and the averages $\langle 3D \rangle_z$ (blue dashed), $\langle 3D \rangle_m$ (purple dotted), $\langle 3D \rangle_{\text{Ross}}$ (solid black), $\langle 3D \rangle_{500}$ (red dashed triple-dotted) and 1D models (solid blue lines).

the above findings with the comparison of the curve of growth. The disk-center intensities are matched best by the averages on column mass density $\langle 3D \rangle_m$, whereas the geometrical averages $\langle 3D \rangle_z$ display the largest discrepancies, in particular for the RG model at solar metallicity with an overestimation by $\sim 60\%$. The results for the averages on optical depth are once again intermediate between the two other kinds of averages. An interesting aspect is the fact that the brightness-temperature averages T_{rad} fail to render the continuum intensities exactly, which has to be interpreted as a consequence of the non-linearity of the Planck function. Our findings are qualitatively similar to those by Uitenbroek & Criscuoli (2011).

5.3. Cautionary remarks

We remind the reader that LTE is often a very poor assumption at low [Fe/H] (e.g. Asplund 2005) and thus that the abundance differences presented in Figs. 7 and 8 should not be added indiscriminately to results from standard 1D LTE abundance analyses. In LTE, the difference between 3D and 1D models can be very substantial for metal-poor stars for especially low excitation and minority species like Fe I (e.g. Asplund et al. 1999; Collet et al. 2007), but those same lines also tend to be sensitive to departures from LTE Bergemann et al. (e.g. 2012); Lind et al. (e.g. 2012) in 1D and $\langle 3D \rangle$ models, mainly due to over-ionization and over-excitation in the presence of a hotter radiation field than the local kinetic temperature (i.e. $J_\lambda > B_\lambda$). Although not explored for more than Li, one would expect that the very cool upper atmospheric layers and thus steep temperature gradients in metal-poor 3D models compared with classical 1D models are even more prone to substantial non-LTE effects (e.g. Asplund et al. 2003; Sbordone et al. 2010). For low excitation FeI lines one would therefore expect the 3D NLTE line strengths to be more similar to the 1D case than the 3D LTE results due to this compensation. We therefore caution the reader

that the here presented 3D LTE abundance corrections (3D LTE - 1D LTE) for Fe I lines are likely to be too negative compared to the NLTE case (3D NLTE - 1D NLTE). As a corollary, it is inappropriate to apply a 1D NLTE abundance correction to a 3D LTE-inferred abundance when the latter is very significant, as often the case at low [Fe/H]. We will return to this important issue in detail in future studies.

5.4. Comparison with 1D models

In Paper I we compared the $\langle 3D \rangle_{\text{Ross}}$ stratifications with 1D models computed with the same EOS and opacity as used in the Stagger-code, in order to quantify the differences arising solely from 1D modeling based on MLT. The 1D models perform quite well at solar metallicity, with the exception of the cool dwarf models (Fig. 7). However, in the metal-poor case the 1D models are not obviously not reproducing correctly the full 3D lines by overestimating the T-stratifications due to the enforcement of radiative equilibrium in the upper atmosphere (Fig. 8). This is in particular distinctive for low excitation neutral iron lines as previously found by Asplund et al. (1999); Collet et al. (2007). Kučinskas et al. (2013) presented similar findings for a solar-metallicity RG simulation as well, namely that the 1D MLT models exhibit lower deviations from the full 3D line formation.

We note that in our 1D models the turbulent pressure is neglected, and the mixing-length is fixed with $\alpha_{\text{MLT}} = 1.5$, both choices that will influence the stratification significantly. Since their effect is strongest in convective zone below the optical surface and the line formation region the influence in terms of abundance is likely small; in fact Kučinskas et al. (2013) found only very small effect $< 0.02\,\text{dex}$ for the reduction in α_{MLT} from 1.5 to 1.2.

6. Conclusions

We have investigated in detail the properties of different methods for computing temporal and horizontal average stratifications from 3D RHD Stagger-grid simulations of stellar surface convection. The choice of the reference depth is critical, as comparisons of the various $\langle 3D \rangle$ demonstrated. We find in general that the temperature stratifications of the $\langle 3D \rangle_z$ and $\langle 3D \rangle_m$ are hotter close to the continuum forming layers and cooler in the upper layers compared to averages on surfaces of constant optical $\langle 3D \rangle_{\mathrm{Ross}}$ and $\langle 3D \rangle_{500}$, while the density shows differences in the opposite sense. The flux-weighted temperature average and brightness temperature average are distinctively hotter than the plain averages, both close the optical surface and in the upper atmosphere, since the Planck function and the fourth power are weighting the larger temperatures more. Averages obtained from the logarithmic values leads to lower temperature and density distributions by giving more weight towards the smaller values in the distribution. These above findings increase with higher T_{eff}, lower $\log g$ and especially with lower [Fe/H].

The statistical properties change depending on the reference depth scale, since the transformation to the new depth scale will inevitably imply a remapping of the values from different heights. The translation to layers of constant optical depth will smooth out temperature fluctuations as a by-product: the temperature is in fact the main source of spatial corrugation of the surfaces of constant optical depth due to the strong temperature sensitivity of the dominant H^- continuum opacity source. Therefore, the temperature contrast and extrema are distinctively reduced, in particular in the superadiabatic region. However, this has also the side-effect that both contrast and minimum-maximum range of the density are enhanced. The concomitant remapping of properties from deeper or higher layers during the transformation to the new reference depth scale will in turn change the average values.

Furthermore, we examined the effects of reversed granulation in the upper layers of metal-poor stars, namely the lowering of temperatures above the granules in metal-poor 3D models compared to classical 1D models. We found that the contribution of radiative re-heating due to weak spectral line absorption features relative to cooling due to mechanical expansion in the upper atmospheric layers is reduced towards higher T_{eff}. On the other hand, the temperature in the regions immediately above the intergranular lanes are primarily controlled by mechanical expansion or compression and appear not to be affected by the reduced metallicity. The two combined effects result in an enhanced contrast in the reversed granulation. This in turn leads to an increase of the corrugation of the surfaces of constant optical depth, which implies that the averages on constant optical depth are sampling values from a very large range in height, thereby affecting the statistical properties.

The comparison of Fe I and Fe II calculated in full 3D and different $\langle 3D \rangle$ atmosphere models reveals the surprising result that the averages on column mass density $\langle 3D \rangle_m$ provide typically the best representation of the 3D model in respect of the line formation. The commonly preferred averages on layers of constant optical depth $\langle 3D \rangle_{\mathrm{Ross}}$ or $\langle 3D \rangle_{500}$ in general perform worse, since the optical depth, $d\tau_\lambda = \rho \kappa_\lambda dz$, contains the additional non-linearity of opacity κ_λ. The differences for the $\langle 3D \rangle_\tau$ are significant in particular for metal-poor models

due to the enhanced reversed granulation in the upper layers. We find the neutral Fe I lines with low excitation potential to feature the largest differences between the mean $\langle 3D \rangle$ and full 3D line calculations. The 1D MLT models perform quite well for the solar metallicity models, however, for metal-poor models the mismatch is evident. Therefore, we caution the use of 1D models for metal-poor stars, which will lead to systematic errors in the spectral analysis. We propose the use of $\langle 3D \rangle_m$ stratifications from the Stagger-grid for spectral line formation purposes.

It is obvious that the temporally and spatially averaged models are incapable of substitute the full 3D atmospheric structure. The reduction due to the averaging will unavoidably lead to the sacrifice of required information. A promising intermediate approach could be the so-called "1.5D" approximation. This approach emulates atmospheric inhomogeneities, which are probed by the traversing radiation, by considering a series of perturbed 1D stratifications for spectral synthesis (e.g. see Ayres et al. 2006). In the spirit of the latter, one could utilize the temporal averaged histograms for an improved spectral line synthesis, since these contain additional information on the statistical distribution of the 3D simulations.

References

Asplund, M. 2005, ARA&A, 43, 481
Asplund, M., Carlsson, M., & Botnen, A. V. 2003, A&A, 399, L31
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Asplund, M., Nordlund, Å., Trampedach, R., Allende Prieto, C., & Stein, R. F. 2000a, A&A, 359, 729
Asplund, M., Nordlund, Å., Trampedach, R., & Stein, R. F. 1999, A&A, 346, L17
Asplund, M., Nordlund, Å., Trampedach, R., & Stein, R. F. 2000b, A&A, 359, 743
Atroshchenko, I. N. & Gadun, A. S. 1994, A&A, 291, 635
Ayres, T. R., Plymate, C., & Keller, C. U. 2006, ApJS, 165, 618
Bensby, T., Johnson, J. A., Cohen, J., et al. 2009, A&A, 499, 737
Bergemann, M., Lind, K., Collet, R., Magic, Z., & Asplund, M. 2012, MNRAS, 427, 27
Böhm-Vitense, E. 1958, ZAp, 46, 108
Cheung, M. C. M., Schüssler, M., & Moreno-Insertis, F. 2007, A&A, 461, 1163
Collet, R., Asplund, M., & Trampedach, R. 2007, A&A, 469, 687
Edvardsson, B., Andersen, J., Gustafsson, B., et al. 1993, A&A, 275, 101
Feautrier, P. 1964, Comptes Rendus Academie des Sciences (serie non specifiee), 258, 3189
Gustafsson, B., Edvardsson, B., Eriksson, K., et al. 2008, A&A, 486, 951
Hayek, W., Asplund, M., Collet, R., & Nordlund, Å. 2011, A&A, 529, A158
Henyey, L., Vardya, M. S., & Bodenheimer, P. 1965, ApJ, 142, 841
Kučinskas, A., Steffen, M., Ludwig, H.-G., et al. 2013, A&A, 549, A14
Lind, K., Bergemann, M., & Asplund, M. 2012, MNRAS, 427, 50
Magic, Z., Collet, R., Asplund, M., et al. 2013, ArXiv e-prints
Mihalas, D., Dappen, W., & Hummer, D. G. 1988, ApJ, 331, 815
Nordlund, A. 1982, A&A, 107, 1
Nordlund, Å. & Stein, R. F. 2001, ApJ, 546, 576
Nordlund, Å., Stein, R. F., & Asplund, M. 2009, Living Reviews in Solar Physics, 6, 2
Pereira, T. M. D., Asplund, M., Collet, R., et al. 2013, ArXiv e-prints
Rutten, R. J., de Wijn, A. G., & Sütterlin, P. 2004, A&A, 416, 333
Sbordone, L., Bonifacio, P., Caffau, E., et al. 2010, A&A, 522, A26
Skartlien, R. 2000, ApJ, 536, 465
Steffen, M. & Holweger, H. 2002, A&A, 387, 258
Steffen, M., Ludwig, H.-G., & Freytag, B. 1995, A&A, 300, 473
Stein, R. F. & Nordlund, A. 1998, ApJ, 499, 914
Uitenbroek, H. & Criscuoli, S. 2011, ApJ, 736, 69

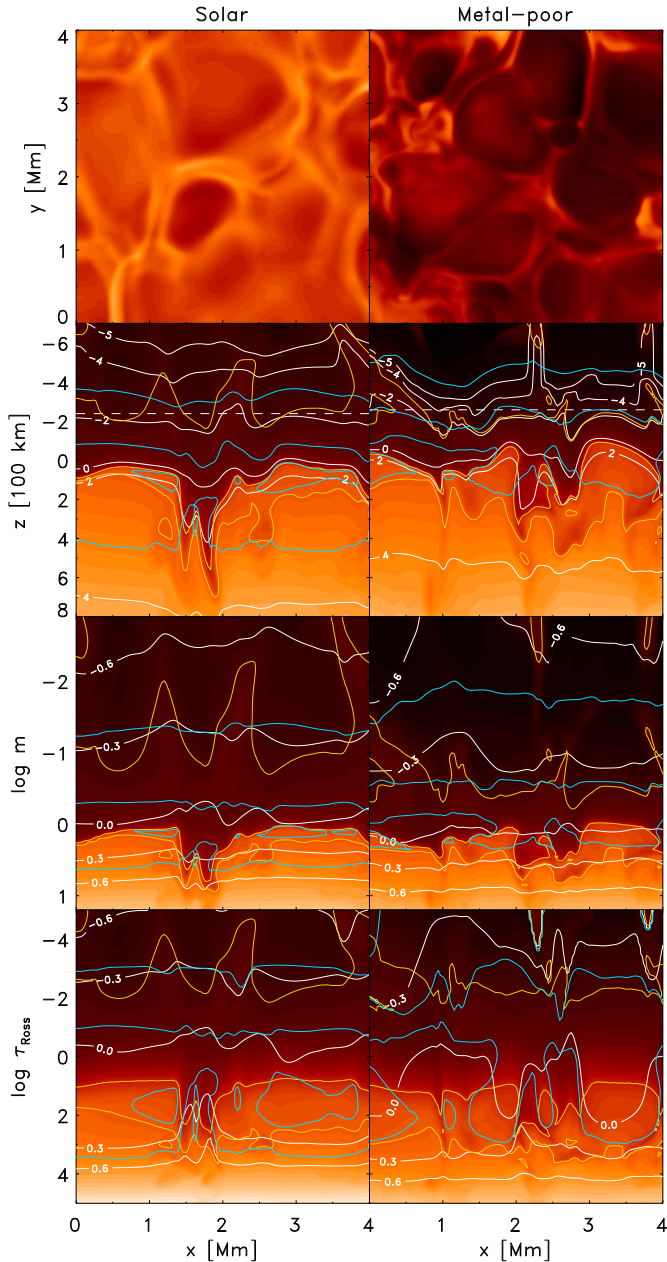**Fig. A.1.** Temperature-contours from our model with $T_{eff} = 6500$ K and $\log g = 4.5$ with [Fe/H] = 0.0 (left) and −3.0 (right). The top panels display horizontal slices with the reversed granulation pattern imprinted in the temperature map (from 3 to 7×10^3 K) taken at ~ 230 km above the surface, which is also indicated in the second panel (dashed lines). The panels below show vertical slices (T-contours from 2 to 17×10^3 K) ranging from $-5.0 \leq \log \tau_{Ross} \leq 5.0$ on layers of constant geometrical height (second), column mass density (third) and Rosseland optical depth (last panel). These panels include isocontours of the temperature (5, 10 and 12×10^3 K; yellow lines) and density (0.1, 1.0 and 2.5×10^{-7} g/cm³; blue lines) and both increase with decreasing vertical depth. We show also lines of constant optical depth (second) and geometrical depth (third and last) indicated with white lines.

Appendix A: Addendum to averaged models

Appendix A.1: Reversed granulation

To illustrate the effects of the remapping of the 3D atmospheric structures onto new reference depth scales, we show in Fig. A.1 slices of temperature contours from our TO-simulation.

We show horizontal temperature maps taken in the upper atmosphere (top panel) and three vertical slices with different reference depth scales, which include geometrical z (second panel), column mass density m (third panel) and Rosseland optical depth (bottom panel). Furthermore, we indicate three different isocontours of the temperature (yellow) and density (blue lines) in Fig. A.1, and we show also lines of constant optical depth τ_{Ross} (white lines in top panel) or geometrical depth z (white lines in middle and bottom panel).

The downdrafts just below the optical surface, denser and cooler than the lighter and hotter surrounding granules, are easily identified (notice the prominent changes T, ρ and τ_{Ross} above the downflows, e.g. $x \approx 1.8$ Mm). Due to the lower temperatures in the downdrafts compared with the granules, the same optical depth value is reached at lower geometrical depths, meaning that the emergent radiation in the intergranular lanes originate from much deeper geometrical heights. Hence, the corrugation of the optical depth on geometrical depth scale is most pronounced in the downdrafts (see isocontour of $\log \tau_{Ross} = 2.0$ in second panel of Fig. A.1).

The opposite is true for the upper atmospheric layers because of the phenomenon of *reversed granulation* (Rutten et al. 2004; Cheung et al. 2007), namely, above the granules, cooling due to adiabatic expansion is dominant, while above the inter granular lanes the radiative reheating and mechanical compression are more important for the energy balance. At lower metallicity and higher T_{eff} the radiative heating above granules is reduced due to the weakening of spectral line features. The resulting reduction in radiative reheating leads to significantly cooler temperatures (see top panel in A.1) and a lower pressure-support, and as a consequence the atmospheric layers at a given constant optical depth subside towards lower geometrical heights, closer to the optical surface. Therefore, the temperature contrast is enhanced in the upper atmosphere. The subsiding of the atmosphere is similar to what we found earlier, namely that the density range spanned in the atmosphere is significantly reduced at lower metallicity (see Fig. 16 in Paper I). Finally, the *enhancement* of the reversed granulation and the temperature contrast results in strongly corrugated surfaces of constant optical depth at the top of metal-poor simulations. We note that we found also an *enhanced* intensity-contrast for metal-poor stars (see Paper I).

The remapping of the individual columns of the 3D structure from geometrical depth to optical depth entails a change of perspective between the old and the new scales in terms of the distribution of values of a particular physical variable at a given constant reference depth. This is again most obvious in the downdrafts in the convection zone (see line of constant geometrical depth at $z = 0.2$ Mm in bottom panel of Fig. A.1). Properties from deeper geometrical heights are mapped onto layers at lower optical depth, and the temperature differences between upflowing and downflowing regions are reduced, which results in smaller temperature contrast and minimum-maximum ranges (see Sect. 4.1). On the other hand, the deviations in the density are significantly *enhanced*, which will clearly alter the statistical properties.

In the upper atmospheric layers of the solar metallicity case, the optical depth is corrugated to only a small amount, therefore the transformation does not affect much the temperature and density (compare the upper flat blue line with the two lower corrugated ones in bottom panel of Fig. A.1). However, the corrugation of the optical depth in the upper atmosphere is
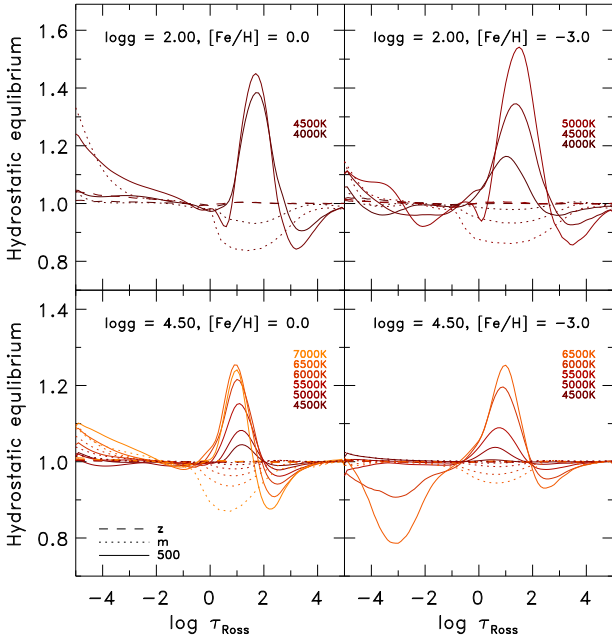
Fig. A.2. Deviations from the hydrostatic equilibrium vs. optical depth. *Dashed lines*: $\langle 3D \rangle_z$ averages; *dotted lines*: $\langle 3D \rangle_m$; *solid lines*: $\langle 3D \rangle_{Ross}$.

rather large for hotter metal-poor stars due to the enhanced reversed granulation. Hence, the effects of remapping onto the optical depth scale for the temperature and density is rather substantial in these simulations. And the distribution of the thermodynamic properties is broadened, such that the meaning of the horizontal average is weakened (see Fig. 6).

In a similar way, the translation to column mass density is reducing naturally the variations in density due to its definition of the reference depth scale, which is the depth integrated density. Therefore, the resulting density fluctuations are rather small in layers at constant column mass density. The variation in temperature is slightly lower than in the averages on geometrical depth, but larger than in the averages on optical depth, as one would expect.

We stress once again that the different reference depth scales are equivalent to each other in terms of the spatial remapping of the 3D atmospheric structures; what differs of course is the statistical properties of physical variables on layers of constant depth, which vary depending on the choice of reference depth scale. In other words, the assignment to a new reference depth scale maps the various physical properties to different depths, thereby altering the statistical properties, meaning that individual properties *appear* differently depending on the reference depth scale. One has to consider two important aspects concerning the horizontal averaging, the first being what kind of quantity is considered, and the second being which reference depth scale is accounted for. Therefore, the statistical properties of the density and temperature are rather distinctive depending on which reference depth scale is considered (see Sect. 4).

Appendix A.2: Hydrostatic equilibrium

The STAGGER-code solves directly the discretized time-dependent radiative-hydrodynamical equations (see Paper I) for the conser-

vation of mass, momentum, and energy. The conservation properties are reflected in the mean $\langle 3D \rangle_z$ stratifications of relaxed, quasi-stationary 3D hydrodynamical models averaged on layers of constant geometrical depth. In particular, the geometrical averages appear over time to be close to hydrostatic equilibrium.[3] To elucidate this further, we analyze the horizontal and time-average of the momentum equation

$$\langle \partial_t \rho \mathbf{v} \rangle = \left\langle -\nabla \cdot (\rho \mathbf{v}\mathbf{v} + \underline{\underline{\tau}}) \right\rangle - \langle \nabla p_{th} \rangle + \langle \rho, \mathbf{g} \rangle \qquad (A.1)$$

with p_{th} being the thermodynamic pressure, $\mathbf{v}$ the velocity field, and $\underline{\underline{\tau}}$ the viscosity stress tensor. Due to the averaging, the only remaining spatial dependence is the vertical one. Divergence terms thus reduce to vertical derivatives, i.e., $\nabla \cdot \langle X \rangle = \partial_z \langle X \rangle$. The time-derivative $\langle \partial_t \rho \mathbf{v} \rangle$ vanishes on time-average as our model atmospheres are relaxed and thus quasi-stationary. The inertial term reduces to turbulent pressure $p_{turb} = \rho v_z^2$; we therefore obtain $\langle \nabla \cdot (\rho \mathbf{v}\mathbf{v}) \rangle = \partial_z \langle p_{turb} \rangle$. The divergence of the viscous stress tensor, $\nabla \cdot \underline{\underline{\tau}}$, vanishes on average. The last two terms yield $\partial_z \langle p_{th} \rangle$ and $\langle \rho g \rangle$, and we retrieve the equation for hydrostatic equilibrium with

$$\partial_z (\langle p_{turb} \rangle + \langle p_{th} \rangle) = - \langle \rho \rangle g. \qquad (A.2)$$

In Fig. A.2 we show the hydrostatic equilibrium in the form of $\rho g dz / dp_{tot} = 1$ for the *temporal* and *geometrical* averaged $\langle 3D \rangle_z$ stratifications, which are very close to hydrostatic equilibrium (dashed lines). We emphasize that the hydrostatic equilibrium is only fulfilled by considering the *total* pressure p_{tot}, as given in Eq. A.2, which includes the non-thermal *turbulent pressure* that occupies a significant fraction of p_{tot} at the top and in the SAR (see Fig. 21 in Paper I).

Furthermore, one can obtain from Fig. A.2 that the averages on a new reference depth scales feature distinctive deviations from hydrostatic equilibrium (see $\langle 3D \rangle_{Ross}$ and $\langle 3D \rangle_m$ with solid and dotted lines, respectively). The transformation of to a new reference depth scale maps all three components of Eq. A.2, geometrical depth z, density ρ and total pressure p_{tot}, away from its hydrostatic equilibrium state. Also, the geometrical depth z looses its strict physical meaning through such a transformation as a mean value. The mean stratifications on constant Rosseland optical depth $\langle 3D \rangle_{Ross}$ deviate slightly at the top and significantly in the SAR from the hydrostatic equilibrium ($\langle 3D \rangle_{500}$ is very similar). The largest departures can be found in the SAR. Furthermore, the amplitude of the discrepancy from hydrostatic equilibrium increases for higher T_{eff} and lower $\log g$.

Appendix A.3: Deviations from the EOS

In 3D RHD simulations, the thermodynamic state of a simulation is self-consistently determined by the EOS. This means in particular that any thermodynamic variable depends on only two independent variables (namely the gas density ρ and the internal energy ε) in a well-defined way. However, the internal self-consistency is broken by reductions like temporal or spatial averaging.

This can be easily understood by investigating the behavior of a function $f(X)$ on a 3D cube of quantity X. For small fluctuations $X' = X - \langle X \rangle$ around the horizontal average at a given depth

---

[3] Note that this statement holds only when considering sufficiently long temporal sequences of snapshots: the individual simulation snapshots at a given instant in time are not in hydrostatic equilibrium.
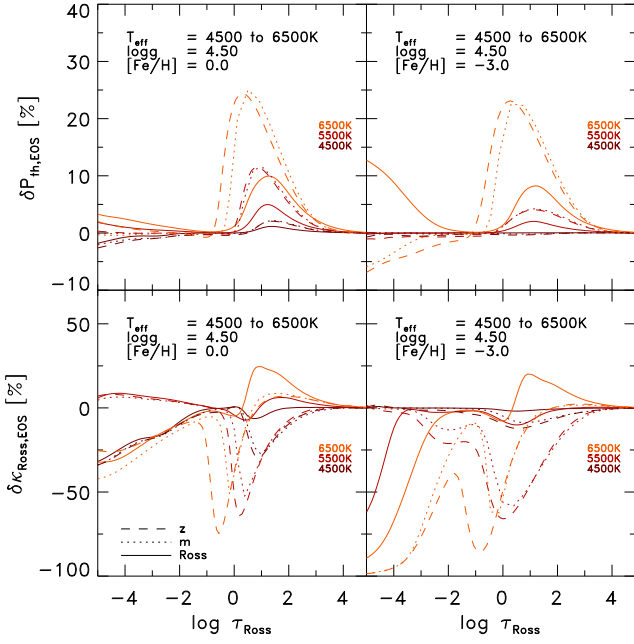
Fig. A.3. Deviations between the spatially and temporally averaged pressure (*top*) and opacity (*bottom*) and the values derived from the EOS, i.e. $X(\langle\rho\rangle, \langle\varepsilon\rangle)$, vs. optical depth. *Dashed lines*: $\langle 3D\rangle_z$ averages; *dotted lines*: $\langle 3D\rangle_m$; *solid lines*: $\langle 3D\rangle_{\mathrm{Ross}}$.

in the model atmosphere, a Taylor-expansion of f up to second order yields

$$f(X) = f(\langle X\rangle + X') \tag{A.3}$$

$$\approx f(\langle X\rangle) + \left.\frac{df}{dX}\right|_{\langle X\rangle} X' + \frac{1}{2}\left.\frac{d^2 f}{dX^2}\right|_{\langle X\rangle} X'^2 \tag{A.4}$$

The horizontal average of this expression evaluates to

$$\langle f(X)\rangle \approx \langle f(\langle X\rangle)\rangle + \left.\frac{df}{dX}\right|_{\langle X\rangle}\langle X'\rangle + \frac{1}{2}\left.\frac{d^2 f}{dX^2}\right|_{\langle X\rangle}\left\langle X'^2\right\rangle \tag{A.5}$$

$$= f(\langle X\rangle) + \left[\frac{1}{2}\frac{d^2 f}{dX^2}\langle X\rangle^2\right]\delta X_{\mathrm{rms}}^2, \tag{A.6}$$

where the definition of the contrast δX_{rms} was used in the last equation (see Eq. 3 in Sect. 4.1). The linear term in Eq. A.5 vanishes as $\langle X'\rangle = 0$ by definition. It is immediately clear that $\langle f(X)\rangle = f(\langle X\rangle)$ holds for linear functions. It is thus the non-linearity of f that causes a departure of $\langle f(X)\rangle$ from $f(\langle X\rangle)$; the departure scales with the square of the contrast δX_{rms}. The discussion can be easily expanded to functions of two variables $f(X, Y)$, as they are found in the EOS.

As a consequence, deriving thermodynamic quantities from averaged independent variables, $\langle\rho\rangle$ and $\langle\varepsilon\rangle$, will lead to inconsistent outcomes. The mean pressure in a given layer of the 3D cube will deviate from the pressure calculated with the EOS from mean density and mean internal energy, $\langle p_{\mathrm{th}}\rangle \neq p_{\mathrm{th}}(\langle\rho\rangle, \langle\varepsilon\rangle)$. Therefore, with $\langle 3D\rangle$ we face another level of complexity.

In order to quantify the deviations, we compute the temperature T, pressure p_{th}, opacity κ_{Ross} and electron number density n_{el} from the EOS by employing the mean independent variables $\langle\rho\rangle$ and $\langle\varepsilon\rangle$. Then, we determine the relative disagreement as $\delta X_{\mathrm{EOS}} = (\bar{X}_{\mathrm{EOS}} - \bar{X})/\bar{X}$. In Fig. A.3, we display the deviations

of thermal pressure $\delta p_{\mathrm{th}}^{\mathrm{EOS}}$ and opacity $\delta\kappa_{\mathrm{Ross}}^{\mathrm{EOS}}$. As suggested by Eq. A.6, we find the maximal deviations typically below the optical surface in the SAR, where the large fluctuations take place due to the overturning, and due the presence of convective motions with their highly asymmetric up and down-flows. Hence, the mean value toddles between the bimodal distribution. Furthermore, we find a strong variation of the δX_{EOS} with stellar parameter, which increases for higher T_{eff} and lower $\log g$. Depending on which reference depth scale is applied the disagreement δX_{EOS} are distinct.

This loss of consistency caused by dimensional reduction means that mean $\langle 3D\rangle$ models can never entirely substitute full 3D models, especially for spectral line formation applications (Uitenbroek & Criscuoli 2011). The mean stratifications are nothing more than statistically meaningful representations of stellar atmospheres, while only the complete 3D data set describes their physical state completely. In 1D model atmospheres, such internal consistency is maintained at all times, since no spatial averaging of non-linear variables is involved in the construction of 1D models.