

3D MODELING OF SOLAR CONVECTION AND ATMOSPHERE DYNAMICS

by

Roar Skartlien

Institute of Theoretical Astrophysics
University of Oslo



Submitted to the Faculty of Mathematics and Natural Sciences,
in partial fulfillment of the requirements for the degree of
Doctor Scientiarum.

July 5, 1998

Acknowledgments

Robert F. Stein and Åke Nordlund are thanked for letting me build onto their convection code. Mats Carlsson is thanked for being my Ph.D. advisor, and for reading the manuscript carefully. Andreas Botnen is thanked for his conversions of the codes to parallel versions running on the SGI CRAY ORIGIN 2000 in Bergen, on the IBM cluster at the University of Oslo and on the working horse “Achernar” at the institute here in Oslo. He is also thanked for help with FORTRAN 90, and for help with solving computer-problems.

Øyvind Andreassen and Per Øyvind Hvidsten at Forsvarets Forskningsinstitutt (FFI) are thanked for letting me use their volume visualization tool VIZ on their fast computers. The group at Colorado Research Associates (CORR) in Boulder, are thanked for their hospitality and for their interest in my subject. Steve Arendt at CORR and Øyvind Andreassen are thanked for introducing me to vortex dynamics. Mark Rast at the HAO in Boulder is thanked for illuminating discussions. Nick Hoekzema at the Sterrekundig Instituut, Utrecht University, is thanked for discussions regarding the real sun versus simulations. Robert J. Rutten, Sterrekundig Instituut, is thanked for help with latex-templates, and for useful writing advice.

Saadatnejad Bard at the institute is thanked for taking me out for coffee and for refreshing walks and baths in the “Nordmarka”-forest, and for astrophysical discussions. Bjørn Håkon Granslo at the institute is thanked for taking me out for astronomical observations that I enjoy in my spare time (which can be in the middle of the week). My other friends are thanked for their understanding when I have to say “not this evening, I have to work”.

A special thank you goes to Teresa Lynne Palmer for her support, and to my parents Bjørg and Asgeir for always supporting my (strange) projects.

This work was also supported by a PhD-grant, and through a grant of computing time (Program for Super-computing), from the Research Council of Norway.

Roar Skartlien,
Institute of Theoretical Astrophysics, Oslo, July 1998.

Contents

1	Introduction	1
1.1	The sun as a stellar laboratory	1
1.2	Scope of the thesis	1
1.3	Solar convection as a wave generator	2
1.4	The dynamic solar atmosphere	4
1.5	Overview of the thesis and summary of the results	7
2	Paper I: Approximate treatment of radiative flux divergence in 3D simulations of stellar atmospheres	11
	Abstract	13
2.1	Introduction	13
2.2	Formulation of the problem	16
2.2.1	LTE opacity and photon destruction probability	16
2.2.2	Source function with coherent isotropic scattering	16
2.2.3	Transfer equation, flux divergence, and lambda operator	17
2.3	Methods	18
2.3.1	The multigroup method	18
2.3.2	The iteration scheme	21
2.3.3	Comparison to other 3D iterative methods	22
2.4	Application to a solar simulation	24
2.4.1	Calculation of group mean opacities	24
2.4.2	Calculation of radiative flux divergence	25
2.5	Test of the methods	28
2.5.1	Monochromatic versus group mean solutions	28
2.5.2	Convergence properties of the iteration method	32
2.6	Discussion	33
2.6.1	Comparison to previous LTE method	33
2.6.2	Solar simulations with and without scattering	34
2.7	Summary	38
3	Paper II: Excitation of wave transients by solar granulation	41
	Abstract	43
3.1	Introduction	43
3.1.1	Turbulent wave excitation in the convection zone	44
3.1.2	Granular wave excitation	44
3.2	Methods	45
3.2.1	Numerical simulation	45
3.2.2	Diagnostics	47
3.3	Results	49
3.3.1	Description of acoustic events	49
3.3.2	Physics of acoustic events	57
3.4	Summary	70

1 Introduction

1.1 The sun as a stellar laboratory

We may look at the solar atmosphere and interior as a laboratory for a wide variety of physical phenomena that is typical for solar-like stars. These include hydrodynamics, magneto-hydrodynamics, thermodynamics and plasma processes. By studying the sun in detail, we learn more about fundamental physics, and we learn more about how stars work. The radiation from the sun gives us more information about these problems than remote stars at such large distances that no telescope today can resolve their surfaces. This, in combination with its influence on the Earth, is why the sun gets so much attention.

It is the atmosphere of the sun that generates the electro-magnetic information we receive here on Earth, or with space observatories. We receive information about the atmosphere itself, but also information about the solar interior. The interior is invisible, and can not be observed directly. Instead we can observe the hydrodynamic effects from the interior on the atmosphere, and learn about the interior by using physical models. Helioseismology has been successful in explaining some aspects of the sun's internal differential rotation, and convective motions. Helioseismology also relies on information contained in radiation escaping from the solar atmosphere.

There are also many unanswered questions regarding the atmospheric physics. A satisfactory understanding of three dimensional dynamics, in terms of flows, waves, magnetic influence, and coupling to radiation is still lacking. Fundamental question are: How is energy transported from the surface of the sun and out to space, how is the atmosphere dynamics driven, and how do magnetic fields behave?

1.2 Scope of the thesis

This work aims at gaining a better understanding of *non-magnetic* hydrodynamics in the solar atmosphere. Specifically, I will study the possibility of generation of acoustic waves in the solar convection zone and/or in the convective overshoot zone, and how these waves can propagate into, and interact with, the overlying atmosphere.

Observations of the solar atmosphere do not reveal the necessary information needed for an unambiguous interpretation of the underlying physics. This is because radiation contains information averaged over large volumes, and physical quantities like temperature and velocity do not translate directly to observables. In addition, wave sources within the convection zone can not be observed directly because radiation can not escape from the convection zone where photons are trapped within the dense plasma. It is only in overlying atmospheric layers that photons can escape and travel to the observer.

To increase the potential for gaining physical insight, I will therefore use three dimensional simulations that include both the convective layers, and the convectively stable atmosphere. These simulations are simplified descriptions of the sun, but they may be good approximations if they produce results in close correspondence to observations.

Simulations of the upper layers of the solar atmosphere have been done by several authors in plane parallel geometry (e.g., Rammacher & Ulmschneider 1992, Carlsson &

Stein 1992, Cheng & Yi 1996 and Carlsson & Stein 1997). These models can not treat convection dynamics, and wave generators due to convection dynamics can therefore not be treated consistently. The current work takes into account all three dimensions with a potential for discovering new phenomena regarding the convection zone - atmosphere relation.

1.3 Solar convection as a wave generator

The convection zone. The outer envelope of the Sun is convective. The solar radius is $R_{\odot}=696\,000$ km, and the depth of the convection zone is about $0.3 R_{\odot}$. Convection is operating because the hydrogen plasma is partially ionized and therefore does not conduct photons very well. The outward energy flux is transported more efficiently by gas motions (advection by convection) rather than photon diffusion, and the energy flux is therefore mainly convective. Hydrogen is fully ionized in convectively stable layers below the convection zone, and the energy flux is here mainly radiative.

The convective flux is exchanged by radiative flux in the uppermost part of the convection zone, where photons escape into space. Upward advection of energy by convection is balanced by radiative cooling. This cooling layer (~ 100 km thick) drives the convection, and is the base of the overlying convectively stable atmosphere. Convective overshoot and turbulent motions here are good candidates for wave sources. Other candidates are turbulent regions within the convection zone.

Insight in solar convection dynamics has improved in recent years due to 3D simulations (e.g., Stein & Nordlund 1989 and Nordlund & Stein 1991). The simulations of Nordlund and Stein show a large correspondence to observations (Spruit et al. 1990). Gas is transported upwards in bright cells (granules) in the cooling layer. After the gas has been radiatively cooled, it flows between these cells to form a dark network (intergranular lanes) with lower temperature than in the cells. The cooled gas descends in these lanes and forms disconnected downdrafts below the cooling layer. These are embedded in upwelling, higher temperature gas.

Wave generation and propagation. A few basic considerations can be made about wave generation and propagation. Acoustic waves can propagate if they have oscillation periods shorter than a certain threshold period. For vertical propagation, this period is the acoustic cutoff period T_{ac} , which is about 3 minutes in the solar atmosphere. Gravity waves can propagate if they have periods longer than a certain threshold period. This is the Brunt-Väisälä period T_{BV} , for almost horizontal propagation. This is also close to 3 minutes, but we have that $T_{BV} > T_{ac}$. These cutoff periods are derived from linear theory for wave propagation, and they may not indicate the proper propagation limits when non-linear effects are important. The cutoff periods are more restricting for inclined wave vectors, with a shorter cutoff period for acoustic waves, and longer cutoff period for gravity waves.

Gravity waves are convectively unstable within the convection zone. They grow in amplitude, and initiate convection. However, they can be generated in the convective overshoot zone in the stable atmosphere, and propagate in the stable surroundings. Evolving

granule topology with typical timescale larger than T_{BV} is needed to produce these waves in the overshoot zone.

Acoustic waves can be generated and propagated within the convection zone, and in the stable atmosphere. Acoustic waves with longer periods than T_{ac} are reflected in the upper convection zone, forming the upper reflecting boundary for the acoustic eigenmodes within the sun (p-modes). Acoustic waves with shorter periods than T_{ac} can propagate from the convection zone and into atmospheric layers without reflection. The typical timescale of the convective motions needs to be shorter than T_{ac} for producing running acoustic waves in the atmosphere. This is also the case for wave sources in the stable, convective overshoot layer. I will concentrate on generation of acoustic waves in this work.

To illustrate acoustic wave sources, I use a quite general equation in pressure with all nonlinear terms included. By combining the conservation equations for mass and momentum in a non-magnetized gas, and the first law of thermodynamics (or the gas energy equation), I get:

$$\left(\frac{\partial^2}{\partial t^2} - c_s^2 \nabla^2 - g \frac{\partial}{\partial z} \right) P = (\Gamma_3 - 1) \left(\frac{\partial \rho q}{\partial t} - g \rho \frac{\partial}{\partial z} \int q dt \right) + c_s^2 \nabla \cdot \vec{F}$$

P is pressure, c_s is local sound speed, g is gravitational acceleration, $\Gamma_3 = (\partial P / \partial e / \rho)_\rho$ and ρ is density. I have neglected the time variation of the local sound speed c_s and Γ_3 for the sake of clarity. q is the net (Eulerian) specific internal energy input rate, which is determined by a balancing between heat input from radiation, viscous dissipation, heat conduction and advection of specific entropy s : $q = Q_{\text{rad.}} + Q_{\text{visc.}} + Q_{\text{cond.}} - \vec{u} \cdot T \nabla s$. $\vec{F} = \nabla \cdot \rho \vec{u} \vec{u} - \vec{F}_v$ is the momentum input rate from the Reynolds stress $\rho \vec{u} \vec{u}$ and from a viscous force \vec{F}_v respectively.

In the context of waves, the left hand side of the equation is the wave operator acting on the pressure fluctuations in P , and the right hand side contains the wave sources. This equation is not only governing waves, but also the background flow field. The different terms have large contributions from the slowly evolving convective flow, in addition to fluctuations that can be associated with waves.

Note that terms which should govern propagating gravity waves in the wave operator are hidden in the right hand side source terms. This is advection of entropy in the vertical direction by adiabatic wave motion (which gives rise to the buoyancy force). Moving these terms into the wave operator gives fourth order time derivatives. For the purpose of illustrating acoustic wave sources, the equation above should be sufficient, and the wave operator here governs propagation of gravity modified acoustic waves. The source terms (from left to right in the equation) represent:

1. Volume changes in time by net heat input. This produces an isotropically propagating pressure disturbance (monopole). A potential source is imbalance between radiative cooling and upward advection of entropy in the cooling layer.
2. Buoyancy forces due to net heat input. The emitted pressure field is dipolar. This can also be associated with the cooling layer.

3. Force divergence. The momentum input rate due to divergence of Reynolds stresses can be interpreted as a force. The divergence of this force creates dilations $\nabla \cdot \vec{u}$, and therefore also pressure fluctuations and sound waves. This term acts as a source in turbulent regions (cf. the theory of Lighthill 1952 regarding jet-noise from turbulence), such as in convective downdrafts within the convection zone. The term can also be expressed in terms of vorticity (Crighton 1981), and is a source of quadrupole sound emission.

These terms have also been discussed elsewhere under more idealized conditions in e.g., Goldreich & Kumar (1990) (plane parallel, isothermal or polytropic stratifications), Rast & Bogdan (1998) (axisymmetric source, polytropic stratification), and without gravity in Howe (1975), and in many textbooks of acoustics.

I study the wave sources for atmospheric acoustic waves in paper II below, using simulation results. Here, I use fully compressible convection-atmosphere simulations in which soundwaves and their sources operate simultaneously and consistently. Also the possible back reaction from waves to sources (and convective flow) are therefore automatically accounted for.

1.4 The dynamic solar atmosphere

Photosphere hydrodynamics. The photosphere extends from the cooling layer, up to 500 km above. The granules (1000-2000 km in diameter for the larger ones) and the intergranular lanes are located at the base of the photosphere, in the cooling layer. The convective flow extends to 500 km above the cooling layer even if these layers are convectively stable. This is the convective overshoot flow, resembling fountains located at each granular cell. Gas cooling by expansion in the divergent overshoot flow is to some degree compensated by heating from radiation coming up from the cooling layer below. Overshooting gas descends into the intergranular lanes, and back into the convection zone. As the gas converges above the intergranular lanes in the photosphere, it compresses and the temperature rises. In this way, we get regions above intergranular lanes with higher temperature than above granules. The temperature contrast in the upper parts of the photosphere is therefore reversed in comparison to the cooling layer, in which the granules are relatively hotter.

The size and evolution of the granules and overshoot flow are controlled by the balance between upward directed convective energy flux, and radiative cooling in the cooling layer. The general evolution pattern is that 1) granules expand, 2) split in several fragments and 3) some fragments diminish in size, while other expand. This behavior is both observed and simulated (Rast et al. 1993 and Rast 1995). The overshoot flow pattern is dictated by the granular cells, hence, evolution in the cooling layer controls the overshoot flow. Granule evolution and rapid changes in the granulation pattern can be potential wave sources.

Atmospheric magnetic fields. Observations from ground or from space (e.g., the Solar Heliospheric Observatory - SOHO) show that the solar atmosphere is penetrated by

magnetic fields (flux tubes) extending up from the convection zone. Since I use pure hydrodynamic simulations in this work, it is of interest to know in which regions of the atmosphere these simulations are applicable.

The plasma conductivity in the sun is large such that the magnetic field lines diffuse relatively slowly through the plasma. The field lines and the gas therefore move together, as if the field lines were frozen into the plasma ¹. Furthermore, photospheric magnetic fields are controlled by the hydrodynamics of convection at the base of the photosphere, since in most cases, the gas pressure here is larger than the magnetic pressure ($\beta > 1$). The motion of the magnetic field in the photosphere appears therefore as advection by horizontally divergent convective flows. The field lines are therefore gathered where the convective flow is convergent, as in intergranular lanes. This effect also tends to form bundles of field lines, the so called flux-tubes. Their footpoints are unevenly distributed across the the photosphere, and the tubes extend into overlying atmospheric layers.

At large scales $\sim 30\,000$ km, the field lines are collected in convergent regions of the supergranular flow, to form the the “magnetic network”. Magnetic fields in the network (and in sunspots) have strengths (flux densities) of a few times 10^3 G leading to $\beta < 1$. The field lines from the network extend into the full range of the overlying atmosphere. Berger et al. (1995) have observed small scale bright points in the photosphere (in the G-band) of typical diameter $\sim 200 - 300$ km, corresponding to network/active regions. These points mark the locations of kilogauss flux tubes (see Berger & Title 1996 and references therein).

Outside the network, in the so called “internetwork”, magnetic fields are weaker. Internetwork fields can be found all over the photosphere and show up as mixed polarity flux on a scale of a few arcseconds. Keller et al. (1994) found that these fields are concentrated in “patches” with strengths less than 10^3 G, or probably even below 500 G. They argued that these patches may be collections of flux tubes with very small cross sections (i.e., too small to be resolved in the observations). These magnetic patches are also believed to advect with the supergranular flow out to the stronger fields in the network region.

We can justify the hydrodynamic simulations used in this thesis in magnetically weak regions ($\beta > 1$). This corresponds to *internetwork* regions in the photosphere and lower chromosphere. Since (weak) internetwork magnetic flux seems to be concentrated to small surface areas (low filling factor), most of the internetwork volume should be described reasonably by hydrodynamics, at least in lower atmospheric layers. In the upper chromosphere, the network flux tubes widen horizontally such that all regions eventually become magnetically dominated ($\beta < 1$).

Chromosphere hydrodynamics. During a total solar eclipse, the chromosphere appears as a “burning forest” above the limb of the moon, with structures changing on timescales of minutes. These structures are finger-like spicules extending up to $\sim 20\,000$ km above the photosphere. Spicules are rooted in the magnetic network, and their dynamics

¹Field lines are not frozen to the plasma at tangential discontinuities where magnetic reconnection operates. Joule dissipation from reconnection is believed to be sufficient to heat the corona to $T \sim 10^6 K$ (Parker 1988). Energy is extracted from convective footpoint-shuffling in the photosphere and released by reconnection in the corona.

is probably strongly influenced by magnetism. The “hydrodynamic region” of the chromosphere is less apparent at first sight. It extends from 500 km to no more than ~ 2000 km above the cooling layer, and is restricted to internetwork regions.

The structure of the hydrodynamic chromosphere has been partially understood by observation and 1D simulations. Observations of “bright grains” and Doppler shifts in the chromospheric CaII H and K spectral lines indicate shocks and non-linear dynamics (see review by Rutten & Uitenbroek 1991). 1D simulations by Carlsson & Stein (1997) have shown that these specific signatures in the CaII H and K lines can be explained by acoustic shocks propagating vertically up from the photosphere. These shocks are modified by non-equilibrium (NLTE) radiation and time dependent hydrogen ionization effects. An excellent correspondence with observations for these simulations has demonstrated that non-magnetic dynamics may be applied in internetwork regions.

These simulations have extremely large temperature fluctuations due to the passage of shocks. The temperature range is about 3000 – 10 000 K at 1000 km height with a mean temperature of about 4500 K (see e.g., Carlsson & Stein 1994). In the sun, these extreme perturbations may be local effects due to local excitation in the convection zone, and not representative of the typical hydrodynamic chromosphere. The “bright grain” simulation of Carlsson & Stein (1997) used a prescribed velocity at the height of 260 km, i.e. well within the photosphere. This velocity was taken from observed Doppler velocities at positions on the sun for which chromospheric bright grains occurred, and hence emphasizing excitation velocities which produce strong shocks. Furthermore, an explanation for why shocks were excited at these selected spatial positions could not be given.

The correspondence between simulated and observed “bright grains”, suggest a large temperature range in the chromosphere. Observations of the infrared molecular spectral lines from CO (carbon monoxide) supports surprisingly low temperatures, with $T < 4000$ K in the lower chromosphere (Ayres et al. 1986). A possibility for coexisting hot and cool components could be explained naturally by large amplitude waves passing through the chromosphere.

Despite the obvious need for simulations that can handle non-linear radiation hydrodynamics, attempts have been made to construct static models to gain insight in the thermal structure (temperature as function of height). The goal for these semi-empirical models has been to reproduce the observed spatially and temporally averaged electro-magnetic spectrum. Gu et al. (1997) has demonstrated that one-component models, such as the VAL-model (Vernazza et al. 1981) can not reproduce the center to limb variations of the radiation temperature in infrared and microwaves. These authors showed that a two-component model, one “hot” and one “cold” component, fits observations better. Ayres et al. (1986) used two-component models to fit CO-line observations and Ca II K lines simultaneously. These two-component models are often referred to as models with “bi-furcated” temperature structures. Since all of the semi-empirical models are static, it is questionable if they can be used to learn about the *solar* chromosphere.

3D simulations are needed to gain further insight in the possible coupling between excitation processes in the convection zone and production of bright grains. This is also needed for a better understanding of the average behavior and thermal structure of the hydrodynamic chromosphere, under the influence of convection dynamics. Important questions

that can be answered by such simulations are:

- How is acoustic wave energy flux produced by convective wave sources?
- How much of this flux is available for chromospheric heating due to shock dissipation?
- How is heat input from shock dissipation balanced by radiative cooling?

1.5 Overview of the thesis and summary of the results

The goal of this work is to explore the effects of convection dynamics on chromospheric dynamics. I concentrate on how acoustic waves are generated by convective flow, and how these waves propagate into, and interact with, the overlying atmosphere.

I use the compressible solar convection model of Nordlund and Stein in its *non-magnetic* form as a basis for this study. An equation of state which accounts for ionization of the most abundant elements on the sun, and radiation cooling of the upper boundary of the convection zone (the cooling layer) is included. This is necessary in order to produce convection dynamics and convection flow topology that represents solar conditions as close as possible. The chromosphere was not included in the Nordlund and Stein model, and the upper boundary was located at 500 km above the cooling layer. I have extended the model to chromospheric layers, to 1500 km above the cooling layer, such that both convection dynamics and chromospheric dynamics can be treated simultaneously. The bottom boundary of the model is located at 1500 km below the cooling layer, within the convection zone.

A large part of this work has consisted of searching the simulation data for atmospheric wave transients that could be linked to convective wave generation. A convective wave source has been detected, and results from this work are given in the following **Paper II**.

Radiation is included in the simulation model since radiative cooling and heating of the gas is important in the energy balance. Radiation was treated in LTE (Local Thermodynamic Equilibrium) in Nordlund and Stein's convection model. This description is sufficient to account for radiative cooling at the top of the convection zone, in the thin cooling layer (thickness of about 100 km that is comparable to one density scale height). This is because the radiation field here is close to the Planck distribution, and because electronic populations are close to thermodynamic equilibrium values. LTE is not sufficient to account for the radiative energy exchange in chromospheric layers, because the radiation field and the electronic population numbers depart from LTE (this is NLTE: Non-Local Thermodynamic Equilibrium). Furthermore, photon scattering is equally important as photon absorption in the chromosphere, and scattering can modify the energy balance considerably in comparison to the LTE description.

In **Paper I**, I have developed an approximate NLTE method that can treat photon scattering and the resulting departure of the radiation field from the Planck distribution. Energy exchange between gas and photons from infrared to ultraviolet wavelengths is accounted for. The full NLTE problem involves millions of atomic transitions, such that approximations are necessary to make the calculations possible on today's computers. The

radiation field is solved for in full 3D geometry, at each timestep in the simulation. It was therefore important to develop this method so as to minimize computer time.

The new radiation method does not modify the energy balance in the cooling layer significantly, and the radiation is close to LTE as for the previous convection model. Hence, the convection dynamics is approximately unaltered, and therefore also the associated convective wave sources. The effect of photon scattering is felt in the upper atmospheric layers, where perturbations in e.g., temperature and velocity are larger than in LTE. The atmospheric response to convective wave sources is therefore influenced by the new radiation method, by giving larger wave amplitudes.

The results of this work can be summarized as follows:

Paper I, Approximate treatment of radiative flux divergence in 3D simulations of stellar atmospheres (Submitted to the journal *Astronomy and Astrophysics*): The approximate NLTE scattering method described in this paper is designed to handle upper atmospheric layers where photon mean free paths are large and photon scattering can be important, and deeper atmospheric layers where photon transport is diffusive. To make this problem computationally tractable, I have made three fundamental approximations: 1) opacities are calculated in LTE, 2) group mean opacities (e.g., Mihalas & Mihalas 1984) are used to describe the opacity spectrum, and 3) coherent and isotropic photon scattering in the source function.

I calculate group mean scattering, absorption and total opacities by wavelength integration of the transport equation. The opacities have different definitions for optically thin and thick layers. The group mean source function is calculated, by the usual definition, as the ratio between emissivity and total opacity. This source function contains an approximate scattering term and an exact contribution from thermal emissivity.

This approach results in a three dimensional scattering problem in each group, which is solved by iteration. I use a one-ray approximation for the mean intensity in the correction procedure, which produces tri-diagonal matrix equations. This requires a minimum of computer time, while still providing relatively fast convergence rate. The CPU-time used for solving for the radiation scales linearly with the number of angles, gridpoints and wavelength groups.

I have compared exact wavelength integrated monochromatic solutions with the corresponding approximate group mean solutions in a snapshot from the simulation model. The total flux divergence obtained from groups deviates with less than 10 % from the exact solution, with maximum deviation in atmospheric layers.

I have also compared two simulations, one with the LTE method used in the convection simulation of Nordlund and Stein (Nordlund 1982 and Nordlund & Dravins 1990), and one with the current scattering method. The difference in atmospheric structure is marginal in the photosphere and in the cooling layer. The temperature fluctuations and wave amplitudes are larger with scattering in chromospheric layers. The maximum chromospheric temperatures are about 1000 K higher, reaching 6500 K. The minimum temperatures in the chromosphere are about the same (2000 K). Higher maximum temperatures result, because heating from viscous dissipation is balanced by radiative cooling at higher temperatures.

Paper II, Excitation of wave transients by solar granulation: I use the radiation method described in paper I in the simulation model. I study wave excitation in the convective flow, and its effect in the overlying atmosphere. I find that transient wave-trains in the photosphere and corresponding shock-trains in the chromosphere are excited by vanishing/collapsing granules. At the time of collapse, upward velocity reverses to downward velocity on a timescale of about 2 minutes. These granules are relatively small and stem from the splitting of larger granules. These events are stochastic exciters for chromospheric dynamics, in the sense that they occur at random times and positions, determined by the chaotic evolution of the granular pattern and convective flow.

Collapsing granules are preferentially located above subsurface downdrafts extending to depths of up to 1000 km below the cooling layer. This suggests that the surface locations of collapsing granules and the associated wave excitation are separated by a typical length scale corresponding to mesogranular scales of the convective flow beneath the surface.

The granule collapse is initiated in the subsurface layers by horizontally converging flow near the downdrafts, on the order of 10 min prior to the flow reversal at the surface. The granule is thereafter subject to an enhanced vertical pressure gradient, and that leads to increasing upflow velocity in the cooling layer. Gravity eventually becomes larger than the pressure gradient, and upflow is reversed to downflow, and a downdraft forms at the site of collapse.

An acoustic wave transient with dominating vertical wave components is excited in the convective overshoot zone (from the cooling layer and to 500 km above this layer) by a negative pressure perturbation ($\sim -30\%$ relative to the unperturbed pressure). The fall- and rise-times of this pressure perturbation are about 2 minutes, i.e., slightly shorter than the 3 min. cutoff period for vertical acoustic wave propagation. The pressure perturbation is mainly driven by density fluctuations (adiabatic changes), but non-adiabatic heating and cooling (heat advection and radiative heating) damps this effect. Several excitation events are detected in the simulation, but only isolated small granules that collapse give a significant response in the form of a decaying atmospheric oscillation.

The transient has dominating amplitudes near 3 min oscillation period, which is the eigen-period for the atmosphere (also the acoustic cut-off period). The envelope of the 3 min oscillation grows from the time of collapse and reaches maximum amplitude 5 minutes afterwards. The vertically propagating wave components steepen, and form shocks in chromospheric layers above $z = 1.0$ Mm. The observed “internetwork bright grains” in the CaII H and K line cores (Rutten & Uitenbroek 1991 and Lites et al. 1993) are produced by vertically propagating acoustic shocks in the chromosphere as shown by the 1D simulations of Carlsson & Stein (1997). Collapsing granules with succeeding shock formation in the chromosphere could therefore be a source of internetwork bright grains.

I find upward directed acoustic wave energy flux (Pu_z -flux) in the atmosphere above the vanishing granules. This is accompanied by a darkening of the granular pattern in continuum intensity. This is supported by the observations of Rimmele et al. (1995), in which upward directed wave energy flux is accompanied by darkening of intergranular lanes.

References

- Ayres T. R., Testerman L., Brault J. W. 1986, *ApJ*, 304, 542
- Berger T. E., Schrijver C. J., Shine R. A., Tarbell T. D., Title A. M., Scharmer G. 1995, *ApJ*, 454, 531
- Berger T. E., Title A. M. 1996, *ApJ*, 463, 365
- Carlsson M., Stein R. F. 1992, *ApJ*, 397, L59
- Carlsson M., Stein R. F. 1994, in M. Carlsson (ed.), *Chromospheric Dynamics*, Proc. Miniworkshop, Inst. Theor. Astrophys., Oslo, p. 47
- Carlsson M., Stein R. F. 1997, *ApJ*, 481, 500
- Cheng Q.-Q., Yi Z. 1996, *A&A*, 313, 971
- Crighton D. G. 1981, *J. Fluid Mech.*, 106, 261
- Goldreich P., Kumar P. 1990, *ApJ*, 363, 694
- Gu Y., Jefferies J. T., Lindsey C., Avrett E. H. 1997, *ApJ*, 484, 960
- Howe M. S. 1975, *J. Fluid Mech.*, 71, 625
- Keller C. U., Deubner F.-L., Egger U., Fleck B., Povel H. P. 1994, *A&A*, 286, 626
- Lighthill M. J. 1952, *Proc. R. Soc. Lond. A*, 211, 564
- Lites B. W., Rutten R. J., Kalkofen W. 1993, *ApJ*, 414, 345
- Mihalas D., Mihalas B. W. 1984, *Foundations of Radiation Hydrodynamics*, Oxford Univ. Press, New York
- Nordlund Å. 1982, *A&A*, 107, 1
- Nordlund Å., Dravins D. 1990, *A&A*, 228, 155
- Nordlund Å., Stein R. F. 1991, in L. Crivellari, I. Hubeny, D. G. Hummer (eds.), *Stellar Atmospheres: Beyond Classical Models*, NATO ASI Series C 341, Kluwer, Dordrecht, p. 263
- Parker E. N. 1988, *ApJ*, 330, 474
- Rammacher W., Ulmschneider P. 1992, *A&A*, 253, 586
- Rast M. P. 1995, *ApJ*, 443, 863
- Rast M. P., Bogdan T. J. 1998, *ApJ*, 496, 527
- Rast M. P., Nordlund A., Stein R. F., Toomre J. 1993, *ApJ*, 408, L53
- Rimmele T. R., Goode P. R., Harold E., Stebbins R. T. 1995, *ApJ*, 444, L119
- Rutten R. J., Uitenbroek H. 1991, *Solar Phys.*, 134, 15
- Spruit H. C., Nordlund Å., Title A. M. 1990, *ARA&A*, 28, 263
- Stein R. F., Nordlund Å. 1989, *ApJ*, 342, L95
- Vernazza J. E., Avrett E. H., Loeser R. 1981, *ApJS*, 45, 635

Paper I

Approximate treatment of radiative flux divergence in 3D simulations of stellar atmospheres

R. Skartlien

Institute of Theoretical Astrophysics, P.O. Box 1029, Blindern, N-0315 Oslo, Norway

Received ; accepted

Abstract. This work presents numerical methods for an approximate solution of the radiative flux divergence in 3D simulations of stellar atmospheres. These methods are designed to handle upper atmospheric layers where photon scattering can be important, and deeper layers where photon transport is diffusive.

By assuming opacity in LTE and coherent isotropic scattering, we calculate group mean opacity coefficients to be used in a group mean source function. This source function contains an approximate scattering term and an exact contribution from thermal emissivity.

The resulting three dimensional scattering problems are solved by iteration using a new method based on a one-ray approximation in the angle integral for the mean intensity. The equations to be iterated are tri-diagonal matrix equations which require a minimum of computer time.

We have compared exact wavelength integrated monochromatic solutions with the corresponding approximate group mean solutions for solar conditions. We find that the total flux divergence obtained from groups deviates with less than 10 % from the exact solution. Flux divergence in individual groups can deviate with typically 30 % in atmospheric layers. When using these groups, the CPU time is reduced by a factor of about 100 in a test case for solar conditions.

Key words: Radiative transfer – Hydrodynamics – Methods: numerical – Sun: atmosphere – Sun: granulation

1. Introduction

Motivation. Radiation is important for the structure and dynamics of stellar atmospheres through its influence on the energy and momentum balance. Any model constructed for gaining insight in stellar atmospheres must include radiation.

Send offprint requests to: R. Skartlien

For the atmosphere of the sun, it is the radiative flux divergence in the energy balance that couples radiation to dynamics, while the radiation force in the momentum balance can be neglected. The upper layer of the solar convection zone is cooled by radiation (positive flux divergence), and convective energy flux is exchanged by radiative energy flux. Gas in atmospheric layers above is influenced by radiative cooling or heating.

The solar convection simulations of Nordlund and Stein (Nordlund 1982 and Nordlund & Stein 1990) have shown an excellent agreement with observed granular structure (Stein & Nordlund 1989, Spruit et al. 1990 and Nordlund & Stein 1991). We would therefore expect that the effect of radiative cooling on convection dynamics in the granular layer has been captured by these models. Radiation was treated in LTE (Local Thermodynamic Equilibrium), with wavelength integrated Planck functions as source functions in a few wavelength groups (Nordlund 1982). This approximation appears to work well in the cooling layer.

These convection simulations covered the photosphere up to typically 500 km above the convective cooling layer. We have extended them to 1500 km, for the purpose of studying connections between convection and chromospheric dynamics.

The chromospheric layers need NLTE (Non-Local Thermodynamic Equilibrium) treatment of radiation. There are millions of atomic transitions that can transfer energy between the “thermal pool” and the radiation field. The associated NLTE radiation hydrodynamics problem is solvable in plane parallel models with present day computers if only a few ions are treated simultaneously (Carlsson & Stein 1992, 1997). In 3D, we need further simplifying assumptions such that the problem becomes computationally tractable. This paper describes the radiation method we use in the extended solar convection-atmosphere simulations, but the same concepts are applicable to simulations of other astrophysical objects.

The first simplifying assumption we make is that the opacity is independent of the radiation field, such that it

can be calculated in thermodynamic equilibrium (from the Boltzmann-Saha equations), and not from the statistical equilibrium equations as in NLTE.

Furthermore, we need to restrict the number of frequency points to a minimum. This can be accomplished by using group mean opacities, which are opacity averages over limited intervals in the spectrum (e.g., Mihalas & Mihalas 1984). This is the second approximation since these opacities do not produce the exact solution of the problem.

A computation of the flux divergence in chromospheric layers where photon scattering is important needs at least the simplest approach, in which we treat the scattering process as coherent and isotropic. Here, photons do not change wavelength and can be scattered in any direction with equal probability in the scattering event. This is the third approximation.

With these three approximations, we can solve isolated scattering problems for each opacity group, with a group mean scattering albedo and a group mean thermal source in a source function in the familiar “two level” format. These scattering problems are solved by iteration, at each timestep of a simulation.

Grey opacities and group mean opacities. A grey opacity can be defined as a wavelength averaged opacity that is a substitute for the full monochromatic opacity spectrum. It is found by integrating the transport equation over wavelength, and from that procedure, defining the average (approximate) opacity which gives the most accurate solution of the integrated intensity. Group mean opacities result from defining mean opacities within limited wavelength intervals (e.g., Mihalas & Mihalas 1984). The more intervals, the more accurate the method is, as we approach the monochromatic solution. The choice of averaging procedure in both of these methods depends on whether the photon transport is diffusive or non-local.

Rosseland (1924) treated the transport equation in the diffusion limit and calculated the well known “Rosseland mean”. This grey opacity comes from wavelength averaging of the series expansion for the intensity up to first order, and assuming that the source function is equal to the Planck function. The Rosseland mean can be viewed as a “mean photon conductivity” (since it is a harmonic mean) rather than a “mean photon resistance” (which is the case for an arithmetically averaged opacity). The Rosseland mean produces the correct radiation flux and flux divergence, as long as the constraint for photon diffusion is satisfied, i.e. that the second and higher derivatives of the intensity on the optical length scale can be neglected.

In optically thin limits, the Planck mean opacity is often used. It has been argued (Pomraning 1971) that the Planck mean is sufficiently accurate since the absorption of photons in optically thin layers is much less than the thermal emission (which is locally determined). Different definitions of mean opacities in optically thick and opti-

cally thin layers leads to the problem of weighting functions between these opacities.

In solar convection simulations, Uus (1987) used three weighting functions for the transition of absorption, emission, and diffusion opacities from optically thin to thick layers. Nordlund & Dravins (1990) used an exponential weighting as a function of Rosseland optical depth. They used the Rosseland mean in deep layers and a mean absorption opacity in optically thin layers. This mean absorption opacity was calculated with the mean intensity from the horizontally averaged model as a weighting function. Only the absorption opacity (total opacity minus scattering opacity) was used to compensate for the non-contribution of scattering in the flux divergence integral. The same opacity was also used in the radiation transport, even if that introduced inaccuracies in the intensities. The full opacity produces more accurate radiation transport.

Finally, Pomraning (1971) developed a scheme more accurate than a simple weighting between grey opacities in optically thin and thick layers. The intensity was expanded in angularly dependent Legendre polynomials, and the solution was found by successive approximations, involving modifications of the scattering kernel in the mean intensity integral in the source function.

The method in this paper uses group averages of the transport equation in which the emissivity term is defined by isotropic and coherent scattering. We define a group mean emissivity consisting of the exact thermal emission term, and an approximate scattering term. In this scattering term, we use the Planck mean scattering opacity in diffusive layers, and the mean intensity averaged scattering opacity in optically thin layers. Here we use the approximate mean intensity of the mean model as the weighting function, as in Nordlund & Dravins (1990). For the optical path-lengths for photon transport, we use the group mean Rosseland opacity in diffusive layers. In optically thin layers, we use the mean intensity averaged total opacity. The ratio between group mean emissivity and group mean opacity defines a group mean source function.

Weighting between thick and thin limits is done using an exponential weighting determined by the mean free photon path set by the group mean Rosseland opacity. Only one free parameter is involved in the weighting between thin and thick limits, namely the maximum mean free path allowed in the diffusion regime. Hence, this weighting is three dimensional, depending on the Rosseland mean free paths at individual gridpoints.

2D and 3D radiative transfer methods. Several methods exist for solving radiative transfer problems for a wide variety of physical problems in two or three dimensions. We could group the methods into (1) the type of atomic physics involved, such as two level or multi level problems, (2) the type of geometry such as 2D slab geometry, 2D axisymmetric problems or full 3D geometry, and (3)

the type of dynamics from pure static cases, to stationary cases, and to the full time dependent problems.

The present method falls in the category of simplified atomic physics in a three dimensional, time dependent medium. In the solar case, the state of the atmosphere changes negligible in the time it takes for a photon to cross it, such that we can neglect the time derivative in the transport equation. We can then solve for the radiation in snapshots from the hydrodynamic part of the code. There are several methods available for this problem, and more general problems involving frequency redistribution in lines. These are reviewed in the following since we use some of their concepts in the solution of our scattering problem.

Cannon developed a multidimensional formalism for his Quadrature Perturbation Techniques. These methods consist of reducing the angle resolution (Cannon 1973a), frequency resolution (Cannon 1973b) or space resolution (Cannon 1976), in approximate operators used in iterative schemes. The full radiation problem can in some cases be set up as a matrix equation where the matrix elements are a numerical representation for the “exact” operator with full quadrature resolution. This matrix equation is computationally demanding, or even impossible to solve on current computers for a 2D or 3D problem. Instead, one uses an iteration series, and solves for the correction to a current estimate of the solution by using the approximate operator.

This principle is widely used, and the methods can be divided into two groups, ALI or AOI depending on the form of operator that is chosen. ALI (Approximate or Accelerated Lambda Iteration) (Cannon 1973b) uses some version of the lambda operator (integral equation formalism), and AOI (Approximate Operator Iteration) uses some version of the differential operator defined by the transfer equation (differential equation formalism).

For all ALI or AOI methods, the iteration is driven by a residual or by an iteration factor (e.g. the variable Eddington-tensor in 2D or 3D, or the Eddington-factor in 1D). These are computed by applying the exact operator on the current estimate of the solution. In most cases, this involves a formal solution for the radiation field given an estimate for the iterated source function. This operation does not use the exact matrix operator directly, but only repeated use of the transfer equation. This is not a prohibitive calculation, even in 3D.

One commonly used method for doing a formal solution in 2D or 3D, is the short characteristics (SC) method (Mihalas et al. 1978 and Kunasz & Auer 1988). The method solves for the intensity along short ray segments between mesh-lines or planes and grid points, using the integral form of the transfer equation.

One alternative method, which is used in this work, takes into account connected rays throughout the medium, rather than short segments. Interpolations of the source function from grid-points to rays are done only in hori-

zontal planes as in Nordlund (1985). The radiation field is solved for along these rays using the second order differential equation of Feautrier (1964).

There are several alternatives for an approximate operator, and the choice is problem dependent. A maximum reduction in spatial coupling (ALI or AOI) is advantageous in 3D. One possibility in ALI is to use the approximate diagonal (local) elements of the lambda operator (Olson et al. 1986) or the exact diagonal elements using the short characteristics method of Olson & Kunasz (1987). Even multilevel 2D or 3D problems become tractable with this approach (Auer et al. 1994). A discussion of approximate local lambda operators and their properties is found in Puls & Herrero (1988). The use of off-diagonal (nonlocal) elements of the lambda operator with arbitrary bandwidth has been studied by Hauschildt et al. (1994) in spherical symmetry. However, using nonlocal elements becomes computationally prohibitive in 2D or 3D.

The SC method with the local lambda operator in ALI-iterations has been used in 2D for the two level CRD (Complete Re-Distribution) problem (Kunasz & Olson 1988). The exact local lambda operator was constructed using the SC formal solver of Kunasz & Auer (1988). The same problem has been studied in 3D geometry by Våth (1994), using short characteristics. Botnen (1997) developed a 3D short characteristics method for periodic horizontal boundary conditions, and included the multilevel package of Carlsson (1986) into that code. Also here, the local lambda operator was used as the approximate operator.

A lambda operator method for the two level CRD line problem with an overlapping continuum in 3D was developed by Nordlund (1985). His method takes advantage of ray decoupling in the mean intensity integral in the source function, such that corrections along individual rays can be calculated independently from other rays. He uses the lambda operator for the bi-directional intensity along rays in this procedure, and not the full angle integrated lambda operator.

Other authors use the differential equation formalism (AOI) rather than the lambda operator (ALI) formalism. Castor et al. (1991) solved the two level CRD problem in 2D using an elliptic differential equation for the mean intensity. They used the first two moments of the transfer equation, which are closed by the Eddington tensor, to arrive at this equation. It was solved by iterating between the Eddington tensor (found by formal solutions using the “exact” angle quadrature) and solutions of the differential equation. A double splitting scheme for the iterations, and a finite element approach for the differential equation were used. Klein et al. (1989) formulated the multilevel CRD or PRD (Partial Re-Distribution) problem in 2D using a finite difference version for the same system of equations. This approach resulted in block tridiagonal matrix equations.

An approximate operator in the AOI scheme has been developed by Cannon (1973a). This consists of approximating the scattering integral in the integro-differential equation that results from inserting the source function with scattering into the transfer equation. Cannon used a reduced angle quadrature in the scattering integral, that makes the solutions for the updated intensities less computationally demanding. Here, the driving error term involved formal solutions with the “exact” angle integral.

Nordlund’s ray decoupling approximation can be viewed as an extreme case of Cannon’s approximation in that the mean intensity integral is reduced to only one ray (two oppositely directed intensity beams). This has a great advantage compared to Cannon’s coupled system of equations that would become prohibitive in 3D.

The new iteration method results when Nordlund’s ray decoupling approximation is applied to Cannon’s iteration scheme on the integro-differential equation. Here, we use Feautrier’s second order differential equation for the bi-directional intensity along rays. The source function includes the isotropic scattering term which involves the angle integral over bi-directional intensity, i.e. we have a set of coupled second order equations. Using the concept of Nordlund (1985) in the angle integral, we decouple these equations and define a correction procedure in an iteration series. The approximate corrections are found by solving tri-diagonal matrix equations along individual rays, and they are thereafter integrated over angles such that the source function can be updated.

The matrix equation for the corrections has simple tri-diagonal structure due to the coherency of the problem. In this way we include nonlocal coupling along rays that gives rapid convergence with a minimum of computations. This is the reason for developing the new iteration method.

Contents of the paper. In Sect. 2 we formulate the monochromatic, coherent and isotropic scattering problem that defines the physical problem to be solved. The multigroup method is developed in Sect. 3, starting from the approach that gives exact solutions in theory, and ending with approximate solutions that are useful in practice. The iteration method is also developed in Sect. 3. Here we use a general notation since the method can be used on both the group mean problem and the monochromatic problem that have the same mathematical form.

We describe how the methods are applied to a solar convection and atmosphere simulation in Sect. 4. Using these simulations, we calculate the errors introduced by using group mean opacities in Sect. 5. Here we also discuss the convergence properties for the iteration method. We compare the group mean opacity method to previous work in Sect. 6, and summarize the results in Sect. 7.

2. Formulation of the problem

In this section we define the monochromatic, coherent and isotropic scattering problem. This defines our fundamental problem that we will solve approximately by using group mean opacities. We also establish our notation in this section.

2.1. LTE opacity and photon destruction probability

We assume that the electronic population numbers are dominated by collisional processes and that we can neglect the influence from the radiation field. In this case we need not solve the rate equations for the populations numbers as for the full NLTE problem, only the Saha-Boltzmann equations. The opacity is in this case therefore only dependent on local quantities, e.g. temperature and density. We denote this approximation as LTE-opacity.

We split the monochromatic opacity χ_λ in an absorption opacity κ_λ that includes all processes that convert photons to thermal energy, and a scattering opacity σ_λ that includes only scattering processes which do not convert photons to thermal energy. The total monochromatic opacity is:

$$\chi_\lambda = \kappa_\lambda + \sigma_\lambda,$$

with units of cross section per volume, $\text{cm}^2 \text{cm}^{-3}$. For notational clarity, we have omitted the general direction dependency resulting from e.g. Doppler shifts of opacity profiles. The scattering and absorption opacities have both contributions from line and continuum processes:

$$\sigma_\lambda = \sigma_\lambda^l + \sigma_\lambda^c,$$

and

$$\kappa_\lambda = \kappa_\lambda^l + \kappa_\lambda^c.$$

We define the photon destruction probability as the probability of absorption once a photon interacts with lines or continua:

$$\epsilon_\lambda = \frac{\kappa_\lambda}{\chi_\lambda} = \frac{\chi_\lambda^l \epsilon_\lambda^l + \kappa_\lambda^c}{\chi_\lambda}.$$

Here, ϵ_λ^l is the photon destruction probability in lines, which is defined as the probability of absorption once a photon interacts with an ionic transition:

$$\epsilon_\lambda^l = \frac{\kappa_\lambda^l}{\chi_\lambda^l},$$

where $\chi_\lambda^l = \kappa_\lambda^l + \sigma_\lambda^l$ is the line opacity only.

2.2. Source function with coherent isotropic scattering

We include nonlocal effects only in the radiation transport (not the opacity) by allowing for scattered radiation in the gas emissivity in addition to thermally generated emissivity. We can write the general angular dependent emissivity as:

$$\eta_\lambda(\Omega) = \int_{\lambda'} \int_{\Omega'} R(\Omega, \Omega'; \lambda, \lambda') I_{\lambda'}(\Omega') d\Omega' d\lambda' + \kappa_\lambda(\Omega) B_\lambda,$$

where Ω is the unit vector along the solid angle Ω , $R(\Omega, \Omega'; \lambda, \lambda')$ is the redistribution function in angle and wavelength, $I_{\lambda'}(\Omega')$ the specific intensity, $\kappa_{\lambda}(\Omega)$ the absorption opacity, and B_{λ} the Planck distribution.

We now assume that the velocity fields are negligible, such that the opacity contributions are direction independent. We further assume that the scattering processes are coherent and isotropic. We note that coherency implies no leakage of photons between different wavelength bands. This can formally be represented by the redistribution function:

$$R(\Omega, \Omega'; \lambda, \lambda') = R(\lambda, \lambda') = \frac{\sigma_{\lambda'}}{4\pi} \delta(\lambda - \lambda'),$$

where $\sigma_{\lambda'}/4\pi$ represent the isotropic (uniform) angular redistribution and the Dirac delta function $\delta(\lambda - \lambda')$ represent coherency (no wavelength shift in the scattering process). With these assumptions, the emissivity becomes:

$$\eta_{\lambda} = \frac{\sigma_{\lambda}}{4\pi} \int_{\Omega'} I_{\lambda}(\Omega') d\Omega' + \kappa_{\lambda} B_{\lambda} = \sigma_{\lambda} J_{\lambda} + \kappa_{\lambda} B_{\lambda}, \quad (1)$$

where J_{λ} is the mean intensity. By definition, the source function is the ratio between emissivity and opacity:

$$S_{\lambda} \equiv \frac{\eta_{\lambda}}{\chi_{\lambda}} = \frac{\sigma_{\lambda}}{\chi_{\lambda}} J_{\lambda} + \frac{\kappa_{\lambda}}{\chi_{\lambda}} B_{\lambda} = (1 - \epsilon_{\lambda}) J_{\lambda} + \epsilon_{\lambda} B_{\lambda}, \quad (2)$$

where we have used the definition of photon destruction probability.

2.3. Transfer equation, flux divergence, and lambda operator

In non-relativistic cases where typical gas velocities are small compared to the speed of light, the radiation field adjusts effectively instantaneously to the evolving medium, and we can use the time independent transport equation to solve for the radiation field:

$$\frac{\partial I_{\lambda}}{\partial r} = \eta_{\lambda} - \chi_{\lambda} I_{\lambda}, \quad (3)$$

where r is the geometrical distance in the direction of I_{λ} . Here we have omitted the angular dependency of I_{λ} for clarity. This is an integro-differential equation since the emissivity is dependent on the angle integrated intensity by Eq. (1).

The radiative flux divergence in the energy equation is obtained from angle and wavelength integration of the transport equation:

$$\nabla \cdot \mathbf{F} = \int_{\lambda} \int_{\Omega} \frac{\partial I_{\lambda}}{\partial r} d\Omega d\lambda = \int_{\lambda} \int_{\Omega} \eta_{\lambda} - \chi_{\lambda} I_{\lambda} d\Omega d\lambda. \quad (4)$$

In our case for isotropic opacity and emissivity, we get, when integrating over solid angle:

$$\nabla \cdot \mathbf{F} = 4\pi \int_{\lambda} \eta_{\lambda} - \chi_{\lambda} J_{\lambda} d\lambda = 4\pi \int_{\lambda} \chi_{\lambda} (S_{\lambda} - J_{\lambda}) d\lambda.$$

By inserting the source function in the last equation, we get:

$$\nabla \cdot \mathbf{F} = 4\pi \int_{\lambda} \epsilon_{\lambda} \chi_{\lambda} (B_{\lambda} - J_{\lambda}) d\lambda.$$

For dominating scattering compared to absorption processes, $\epsilon_{\lambda} \sim 0$ and radiation couples only weakly to the gas and the contribution to the energy balance is less important. When the gas temperature is larger than the radiation temperature, such that $B_{\lambda} > J_{\lambda}$, more photon energy is emitted than absorbed, and we get positive flux divergence or radiative cooling of the gas. And vice versa, net absorption of photons occurs if the gas temperature is lower than the radiation temperature. In media where the photon mean free path is relatively large compared to the length scale of temperature fluctuations, J_{λ} is smoother than B_{λ} and radiation serves to smooth out temperature perturbations in the gas. For high density, optically thick media we have $B_{\lambda} \sim J_{\lambda}$ and $\epsilon_{\lambda} \sim 1$, which again gives little contribution from the flux divergence in the energy balance.

We define, for later use, the lambda operator for the mean intensity in terms of the volume integral of the source function and the kernel function K . Using the transport equation, it can be shown (Rybicki 1984) that the mean intensity is given by:

$$J_{\lambda}(\mathbf{r}) = \Lambda(\chi_{\lambda})[S_{\lambda}] = \int_V S_{\lambda}(\mathbf{r}') K_{\lambda}(\mathbf{r}, \mathbf{r}') d^3 r',$$

where the kernel function that defines the Λ -operator is given by:

$$K_{\lambda}(\mathbf{r}, \mathbf{r}') = \frac{\chi_{\lambda}(\mathbf{r}')}{4\pi r^2} e^{-\Delta\tau_{\lambda}}, \quad (5)$$

where $r = |\mathbf{r} - \mathbf{r}'|$ is the distance from the point \mathbf{r} where the mean intensity is evaluated to an arbitrary point \mathbf{r}' . The corresponding optical path length between these points is given by:

$$\Delta\tau_{\lambda} = \int_0^r \chi_{\lambda}(\mathbf{r}') dr. \quad (6)$$

The kernel is geometrically small for optically thick media, as seen by the exponential dependency on the optical path length, and the mean intensity is approximately equal to the local source function. For larger optical path lengths, the kernel spans larger volumes and the mean intensity is non-locally determined, and smoother than the source function.

We can formulate the radiation problem as an integral equation in the source function (Fredholm equation of the second kind) by using our definition of the source function in Eq. (2), and the lambda operator:

$$S_{\lambda} = (1 - \epsilon_{\lambda}) \Lambda(\chi_{\lambda})[S_{\lambda}] + \epsilon_{\lambda} B_{\lambda}. \quad (7)$$

This equation is linear since the opacity used in the lambda operator is independent of the radiation field (the solution). In principle, we could solve this equation numerically by representing the lambda operator as a matrix. The matrix would consist of the discretized kernel function, and include N^2 elements where N is the number of gridpoints. A serious problem arises in three dimensions. For a simulation with 100 gridpoints in each direction, we would have to deal with $N = 10^6$ unknowns

and $N^2 = 10^{12}$ matrix elements. The intractable size of the matrix is the reason for using iterative methods.

We note that a large amount of wavelength points are needed for an accurate evaluation of the radiative flux divergence in Eq. (4). Each wavelength point needs a solution of a problem equivalent to the integral equation, and it is prohibitive to solve this scattering problem at several hundred wavelength points per timestep in simulations. The multigroup method described below is designed to solve this problem approximately by using a limited set of wavelength groups. The scattering problem is solved within each group, and an approximate radiative flux divergence is found by adding the group-contributions.

3. Methods

First, we develop the multigroup method, then the iteration scheme, and finally we compare the new iteration method with existing methods. We also show the advantages of the new iteration method for this specific scattering problem.

3.1. The multigroup method

Determination of groups. We follow the group definition of Nordlund (1982). The atmospheric height range is divided into a chosen number of intervals. A set of wavelengths belong to the same group if the associated monochromatic optical depths are unity within the same height interval. This definition serves to sort wavelength points with similar shapes of the kernel function into the same group. Furthermore, strong lines fall within one group corresponding to high layers, and the continua fall within one group corresponding to deeper layers.

Specifically, the height scale is chosen as a standard optical depth along the vertical direction rather than geometrical height. The standard and monochromatic optical depths can be calculated in the plane parallel atmosphere of the mean model (horizontally and temporally averaged). The group number i for monochromatic wavelength j is determined by:

$$i = \text{Int} \left(\frac{\log(\tau_0(\tau_j = 1))}{\Delta \log(\tau)} \right) + \text{const.},$$

where Int means nearest integer value, $\tau_0(\tau_j = 1)$ is the standard optical depth where the monochromatic optical depth is unity, and $\Delta \log(\tau)$ depends on the number of groups chosen.

We will calculate integrals over wavelength within each group in the following. The set of wavelengths λ_j that belongs to group i is denoted by $\Delta \lambda_i$. This set need not be continuous, since wavelengths in infrared and ultraviolet can be included in the same group, while a wavelength in the visible can be included in another group. An integration within each group can be carried out also in this case.

This is implemented numerically as a quadrature sum:

$$Y_i = \sum_{\Delta \lambda_i} Y_{\lambda_j} w_j,$$

where λ_j and w_j are the wavelength quadrature points and weights respectively, and Y_{λ_j} an arbitrary wavelength dependent quantity. In the following, we denote this quadrature sum by the notation of continuous integrals:

$$Y_i = \int_{\Delta \lambda_i} Y_{\lambda} d\lambda.$$

Exact solution. Here we derive the exact transport equation for the wavelength integrated intensity within a wavelength group. The result can not be used in practice to solve for the wavelength integrated intensity, since it involves knowledge of the monochromatic radiation field which we will not solve for. The derivation is made to arrive at approximate solutions that can be used without having knowledge of the monochromatic radiation field. Integration of the monochromatic transport equation (Eq. 3) over the subset of wavelengths $\Delta \lambda_i$ gives:

$$\int_{\Delta \lambda_i} \frac{\partial I_{\lambda}}{\partial r} d\lambda = \int_{\Delta \lambda_i} (\sigma_{\lambda} J_{\lambda} + \kappa_{\lambda} B_{\lambda} - \chi_{\lambda} I_{\lambda}) d\lambda.$$

Here we have used Eq. (1) for the monochromatic emissivity. We use the following notation for the integrated transport equation:

$$\frac{\partial I_i}{\partial r} = \sigma_i^J J_i + \kappa_i^B B_i - \chi_i^I I_i. \quad (8)$$

The integrated intensity is:

$$I_i = \int_{\Delta \lambda_i} I_{\lambda} d\lambda,$$

the integrated mean intensity is:

$$\begin{aligned} J_i &= \int_{\Delta \lambda_i} J_{\lambda} d\lambda \\ &= \frac{1}{4\pi} \int_{\Delta \lambda_i} \int_{\Omega} I_{\lambda} d\Omega d\lambda = \frac{1}{4\pi} \int_{\Omega} I_i d\Omega, \end{aligned}$$

and the integrated Planck distribution is:

$$B_i = \int_{\Delta \lambda_i} B_{\lambda} d\lambda.$$

The following definitions satisfy Eq. (8) for I_i exactly in combination with J_i and B_i in the emissivity term. The group mean scattering opacity is:

$$\sigma_i^J = \frac{\int_{\Delta \lambda_i} \sigma_{\lambda} J_{\lambda} d\lambda}{J_i}, \quad (9)$$

which is dependent on the unknown J_{λ} . The Planck mean absorption opacity is:

$$\kappa_i^B = \frac{\int_{\Delta \lambda_i} \kappa_{\lambda} B_{\lambda} d\lambda}{B_i}. \quad (10)$$

The group mean opacity that defines the mean optical path lengths for Eq. (8) is:

$$\chi_i^I = \frac{\int_{\Delta \lambda_i} \chi_{\lambda} I_{\lambda} d\lambda}{I_i}. \quad (11)$$

This opacity is direction dependent due to the direction dependency of I_λ , and furthermore it is dependent on the unknown I_λ .

The group mean source function for this problem follows from the ratio between group mean emissivity and group mean opacity from Eq. (8). Hence,

$$S_i = \frac{\sigma_i^J J_i + \kappa_i^B B_i}{\chi_i^I}.$$

Eq. (8) results in a lambda operator for J_i whose kernel is a function of χ_i^I only (analog to Eq. 5), and that acts on S_i :

$$J_i = \Lambda(\chi_i^I)[S_i].$$

Approximate solution in the streaming regime. In regions where radiation is non-diffusive (streaming regime) we have to solve the transport equation to account for non-local radiation coupling within the gas. The scattering opacity σ_i^J (Eq. 9) and total opacity χ_i^I (Eq. 11) in the exact transport equation (Eq. 8), can not be pre-calculated without knowledge of the monochromatic intensity I_λ . Instead, we use the monochromatic radiation field from the same scattering problem from the plane parallel version of the atmosphere. This is a one-time computation and it is also relatively fast. We get the one dimensional intensity I_λ^{pp} and mean intensity J_λ^{pp} , where label ‘‘pp’’ denotes ‘‘plane parallel’’. Both of these quantities are a function of atmospheric height.

As an approximation, we solve Eq. (8) with scattering opacity and total opacity that are pre-calculated with this plane parallel radiation field. We assume that $J_\lambda \simeq J_\lambda^{\text{pp}}$, and replace the scattering opacity σ_i^J with the approximate σ_i^{Jpp} :

$$\sigma_i^J \simeq \sigma_i^{\text{Jpp}} = \frac{\int_{\Delta\lambda_i} \sigma_\lambda J_\lambda^{\text{pp}} d\lambda}{\int_{\Delta\lambda_i} J_\lambda^{\text{pp}} d\lambda}.$$

Similarly, we could choose to use I_λ^{pp} in the calculation of the total opacity. This would, however, introduce a direction dependency. As a first approximation we eliminate the direction dependency by setting $I_\lambda \simeq J_\lambda^{\text{pp}}$ in the averaging, and replace χ_i^I with χ_i^{Jpp} :

$$\chi_i^I \simeq \chi_i^{\text{Jpp}} = \frac{\int_{\Delta\lambda_i} \chi_\lambda J_\lambda^{\text{pp}} d\lambda}{\int_{\Delta\lambda_i} J_\lambda^{\text{pp}} d\lambda}. \quad (12)$$

Hence, the approximate wavelength averaged intensity I_i^* is defined by the transport equation:

$$\frac{\partial I_i^*}{\partial r} \equiv \sigma_i^{\text{Jpp}} J_i^* + \kappa_i^B B_i - \chi_i^{\text{Jpp}} I_i^*, \quad (13)$$

where the approximate mean intensity is:

$$J_i^* = \frac{1}{4\pi} \int_{\Omega} I_i^* d\Omega.$$

The ratio between emissivity and opacity gives the source function:

$$S_i^* = \frac{\sigma_i^{\text{Jpp}} J_i^* + \kappa_i^B B_i}{\chi_i^{\text{Jpp}}}. \quad (14)$$

The Planck mean absorption opacity in Eq. (10) remains unchanged, since the Planck distribution is always defined for a given temperature. Hence, the thermal emissivity $\kappa_i^B B_i$ has remained exact. Since Eq. (13) has the standard transport equation format, we get a lambda operator kernel that is a function of χ_i^{Jpp} , and that acts on the source function S_i^* :

$$J_i^* = \Lambda(\chi_i^{\text{Jpp}})[S_i^*]. \quad (15)$$

The largest inaccuracies in the approach given above result from using J_λ^{pp} as a weight function in the averaging of the total opacity (Eq. 12), since the correct weight function I_λ is strongly direction dependent. The effect in J_i^* compared to the exact J_i can be predicted by simple arguments for the stellar atmosphere case.

χ_i^I is higher for incoming rays compared to outgoing rays. This can be understood by studying the integrand in Eq. (11). Incoming intensity is characterized by emission lines that favor weighting of opacity peaks in lines, while the outgoing intensity is characterized by absorption lines that favor weighting of the continua. Hence, for isotropic χ_λ , χ_i^I is anisotropic due to the direction dependent spectrum of I_λ .

Since J_λ^{pp} is angle independent and intermediate between incoming and outgoing intensities, χ_i^{Jpp} will have an intermediate value between the values for χ_i^I that correspond to incoming and outgoing rays. In result, the kernel function for χ_i^I decays more slowly towards the photosphere (kernel for outgoing rays) than the kernel for χ_i^{Jpp} . And vice versa, the kernel function for χ_i^I decays more rapidly toward the upper boundary (kernel for incoming rays) than the kernel for χ_i^{Jpp} .

This gives, by using the Eddington-Barbier relation and an outward decreasing source function, an I_i^* that is lower than the exact I_i for both incoming and outgoing intensities. This gives finally a lower J_i^* than the exact J_i . We will study this deviation in Sect. 5.

Approximate solution in the diffusion limit. The diffusion limit is characterized by small photon mean free paths (high opacity) compared to the scale of variation of the source function. In this case, the lambda operator is local due to the rapid cutoff of the kernel function by the exponential dependency of the optical path length (Eq. 5) that is large for small geometrical distances (Eq. 6). Hence, $J_\lambda \simeq S_\lambda$. This implies $S_\lambda \simeq B_\lambda$ by using the source function (Eq. 2), even if scattering is dominating. By expanding the source function in a power series, it can be shown (e.g., Mihalas 1978) that the solution of the transport equation (Eq. 3) is to first order (using $S_\lambda = B_\lambda$):

$$I_\lambda = B_\lambda - \frac{1}{\chi_\lambda} \frac{\partial B_\lambda}{\partial r}.$$

The higher order terms can be neglected in the diffusion limit. The wavelength averaged intensity is given by:

$$I_i = B_i - \frac{1}{\chi_i^R} \frac{\partial B_i}{\partial r}, \quad (16)$$

where χ_i^R is the Rosseland group mean opacity defined by:

$$\frac{1}{\chi_i^R} = \int_{\Delta\lambda_i} \frac{1}{\chi_\lambda} \frac{\partial B_\lambda}{\partial r} d\lambda \bigg/ \int_{\Delta\lambda_i} \frac{\partial B_\lambda}{\partial r} d\lambda.$$

We want to use a form of the transport equation in both streaming and diffusion limits such that both limits will be satisfied in a global numerical solution. The alternative would be to split the numerical scheme, and compute the Rosseland flux divergence directly by using 3D gradient operators in diffusive layers and use the transport equation in the streaming limit in atmospheric layers. The transport equation that satisfies Eq. (16) in the diffusion limit, results from using χ_i^R as the total opacity:

$$\frac{1}{\chi_i^R} \frac{\partial I_i^*}{\partial r} = S_i^* - I_i^*, \quad (17)$$

where we must assure that $S_i^* \rightarrow B_i$. Again,

$$J_i^* = \Lambda(\chi_i^R)[S_i^*]. \quad (18)$$

For small mean free paths, we have $J_i^* \simeq S_i^*$.

The correct weight used in the source function for the scattering and total opacities is now the Planck distribution since the mean intensity approaches the Planck distribution in the limit. This also gives the correct 3D variations, which the plane parallel mean intensity can not account for. Hence, Eq. (14) is rewritten as:

$$S_i^* = \frac{\sigma_i^B J_i^* + \kappa_i^B B_i}{\chi_i^B}, \quad (19)$$

where the group mean scattering opacity is given by:

$$\sigma_i^B = \frac{\int_{\Delta\lambda_i} \sigma_\lambda B_\lambda d\lambda}{B_i}, \quad (20)$$

and the Planck mean opacity by:

$$\chi_i^B = \frac{\int_{\Delta\lambda_i} \chi_\lambda B_\lambda d\lambda}{B_i}. \quad (21)$$

As $J_i^* \rightarrow S_i^*$, this source function satisfies the requirement $S_i^* \rightarrow B_i$. It also satisfies $S_i^* = B_i$ for pure absorption where $\kappa_i^B = \chi_i^B$, and $S_i^* = J_i^*$ for pure scattering where $\sigma_i^B = \chi_i^B$.

Weighting between diffusion and streaming regimes. The source function for the streaming limit in Eq. (14) and for the diffusion limit Eq. (19) is combined into one source function using a weighting function between the two limits:

$$S_i^* = \left(\frac{\sigma_i^{\text{JPP}} \rightarrow \sigma_i^B}{\chi_i^{\text{JPP}} \rightarrow \chi_i^B} \right) J_i^* + \frac{\kappa_i^B B_i}{\chi_i^{\text{JPP}} \rightarrow \chi_i^B} \equiv \rho_i J_i^* + \eta_i, \quad (22)$$

where we have defined a group mean scattering albedo ρ_i and a group mean thermal source η_i . The arrows indicate transition to the diffusion limit in optically thick layers. We use the same weighting function for the group mean opacities used in the lambda operators in Eq. (15) and Eq. (18):

$$\chi_i^{\text{JPP}} \rightarrow \chi_i^R \equiv \chi_i^T, \quad (23)$$

where we have defined the combined group mean opacity χ_i^T . The mean intensity for both regimes is given by:

$$J_i^* = \Lambda(\chi_i^T)[S_i^*],$$

or in terms of the transport equation:

$$\frac{\partial I_i^*}{\partial r} = \chi_i^T (S_i^* - I_i^*). \quad (24)$$

What remains to define is the weighting function. We treat the radiation problem as non-diffusive within a group when the Rosseland photon mean free path $l_i^{\text{mfp}} = 1/\chi_i^R$ is larger than a typical length scale of variation l_i of the group mean Planck function. This weighting is three dimensional since the Rosseland group mean opacity varies from gridpoint to gridpoint. We choose the ‘‘optically thin weight’’ by the exponential:

$$W_{\text{stream}} = e^{-l_i \chi_i^R}.$$

For weighting between streaming regime opacities X_{stream} and diffusion regime opacities X_{diffuse} in both source function and transport equation, we use:

$$X = W_{\text{stream}} X_{\text{stream}} + (1 - W_{\text{stream}}) X_{\text{diffuse}}, \quad (25)$$

which in our arrow notation is:

$$X = X_{\text{stream}} \rightarrow X_{\text{diffuse}}.$$

Consistency in the monochromatic limit. We recover the exact monochromatic solution if the wavelength interval of a group approaches zero. In this limit, the form of the weighting functions in the opacity averages plays no role, and the source function in Eq. (22) becomes the monochromatic source function in Eq. (2). The opacity used in the transfer equation, Eq. (23), becomes the monochromatic opacity. The form of depth weighting plays no role in this case, since the transition is made between two identical quantities.

Approximate flux divergence. The exact wavelength integrated flux divergence is for each group:

$$\nabla \cdot \mathbf{F}_i = \int_{\Omega} \int_{\Delta\lambda_i} \frac{\partial I_\lambda}{\partial r} d\lambda d\Omega = \int_{\Omega} \frac{\partial I_i}{\partial r} d\Omega. \quad (26)$$

For the approximate I_i^* from Eq. (24), we get the flux divergence:

$$\nabla \cdot \mathbf{F}_i^* = \int_{\Omega} \frac{\partial I_i^*}{\partial r} d\Omega = \int_{\Omega} \chi_i^T (S_i^* - I_i^*) d\Omega. \quad (27)$$

This becomes in the streaming limit, using Eq. (14) and Eq. (13):

$$\int_{\Omega} \chi_i^{\text{JPP}} (S_i^* - I_i^*) d\Omega = 4\pi (\kappa_i^B B_i - \kappa_i^{\text{JPP}} J_i^*),$$

where $\kappa_i^{\text{JPP}} = \chi_i^{\text{JPP}} - \sigma_i^{\text{JPP}}$ is the group mean absorption opacity. We note that the approximation lies in the absorption term $\kappa_i^{\text{JPP}} J_i^*$, and that the thermal emission term $\kappa_i^B B_i$ is exact. The approximate flux divergence in the diffusion limit becomes, when using Eq. (16), Eq. (17) and

the source function in Eq. (19) in the limit $S_i^* \rightarrow B_i$:

$$\int_{\Omega} \chi_i^R (B_i - I_i^*) d\Omega = - \int_{\Omega} \frac{\partial}{\partial r} \frac{1}{\chi_i^R} \frac{\partial B_i}{\partial r} d\Omega.$$

The total exact flux divergence is obtained by summation over all groups according to Eq. (4) and Eq. (26):

$$\nabla \cdot \mathbf{F} = \sum_i \nabla \cdot \mathbf{F}_i = \int_{\Omega} \int_0^{\infty} \frac{\partial I_{\lambda}}{\partial r} d\lambda d\Omega,$$

and similarly, we have that the total approximate flux divergence is:

$$\nabla \cdot \mathbf{F}^* = \sum_i \nabla \cdot \mathbf{F}_i^*. \quad (28)$$

3.2. The iteration scheme

Notation definition. Here, we derive the iteration method to solve for either the monochromatic radiation or for the radiation in individual groups. Both problems have the same format for the transport equation and we will therefore use a general formalism in this section. The general transport equation is:

$$\pm \frac{1}{\chi} \frac{\partial I^{\pm}}{\partial r} = S - I^{\pm}, \quad (29)$$

where χ is χ_i^T for groups, and χ_{λ} for the monochromatic problem. I^{\pm} is the integrated intensity for groups, or the monochromatic intensity, in the two opposite directions along a given ray. This ray has an arbitrary direction, and r is the geometrical distance along this ray from an arbitrary point. I^+ is chosen to be in the direction of increasing distance r and I^- in the direction of decreasing r . The source function S is given by Eq. (22) for groups, and by Eq. (2) for the monochromatic problem. In both cases, we write S in terms of a scattering albedo ρ , mean intensity J and thermal source η :

$$S = \rho J + \eta. \quad (30)$$

The iterative method uses the bi-directional average intensity as the unknown variable:

$$p_{\Omega} = \frac{1}{2}(I^+ + I^-),$$

where Ω denotes the direction dependency. The transfer equations for I^+ and I^- (Eq. 29) can be recast to the familiar second order differential equation first formulated by Feautrier (1964):

$$\left(1 - \frac{\partial^2}{\partial \tau_{\Omega}^2}\right) p_{\Omega} = S. \quad (31)$$

τ_{Ω} is the optical path-length along the associated ray, and the relation to opacity and geometrical distance is:

$$d\tau_{\Omega} = \chi dr.$$

We define the inverse lambda operator for a single ray as:

$$\Lambda_{\Omega}^{-1} \equiv 1 - \frac{\partial^2}{\partial \tau_{\Omega}^2}. \quad (32)$$

Hence, p_{Ω} can be written in terms of the lambda operator for a single ray:

$$p_{\Omega} = \Lambda_{\Omega}[S].$$

The explicit form of this operator follows from using the integrated transport equation (Eq. 29) for I^+ and I^- :

$$\Lambda_{\Omega}[S] = \frac{1}{2} \left(\int_0^{\infty} S e^{-\tau_{\Omega}^+} d\tau_{\Omega}^+ + \int_0^{\infty} S e^{-\tau_{\Omega}^-} d\tau_{\Omega}^- \right), \quad (33)$$

where τ_{Ω}^+ and τ_{Ω}^- are the optical lengths measured along opposite directions for a given ray. Note that p_{Ω} can be found in two ways for a given source function; either by solving the differential equation $\Lambda_{\Omega}^{-1}[p_{\Omega}] = S$, or by integrating over the source function directly, using $p_{\Omega} = \Lambda_{\Omega}[S]$. The mean intensity is given by:

$$J = \int_{\Omega} p_{\Omega} \frac{d\Omega}{2\pi}. \quad (34)$$

Hence, the full lambda operator is formally given by the relations:

$$J = \Lambda[S] = \int_{\Omega} \Lambda_{\Omega}[S] \frac{d\Omega}{2\pi}. \quad (35)$$

The integration is done over a half-sphere (spanning the solid angle 2π). By combining Eq. (34), Eq. (32) and Eq. (31) with Eq. (30), we get the integro-differential equation for p_{Ω} that specifies our problem:

$$\Lambda_{\Omega}^{-1}[p_{\Omega}] = \rho \int_{\Omega} p_{\Omega} \frac{d\Omega}{2\pi} + \eta. \quad (36)$$

We can alternatively formulate the problem as an integral equation in the source function by combining Eq. (35) with Eq. (30):

$$S = \rho \Lambda[S] + \eta. \quad (37)$$

We use the integro-differential form in the derivation of the iteration method.

Iterations with ray decoupling. With the formalism given above, we are in a position to derive the iteration method. Let p_{Ω}^n be the current estimate of the unknown p_{Ω} at iteration step n . The corresponding estimate of the source function is, by using Eq. (30) and Eq. (34):

$$S^n = \rho \int_{\Omega} p_{\Omega}^n \frac{d\Omega}{2\pi} + \eta. \quad (38)$$

The corresponding formal solution is given by:

$$p_{\Omega}^{\text{FS}} = \Lambda_{\Omega}[S^n], \quad (39)$$

and let S^{FS} be given by:

$$S^{\text{FS}} = \rho \int_{\Omega} p_{\Omega}^{\text{FS}} \frac{d\Omega}{2\pi} + \eta, \quad (40)$$

which is equivalent to $S^{\text{FS}} = \rho \Lambda[S^n] + \eta$. Since the solution S satisfies Eq. (37), we have that both S^{FS} and S^n approach S for a convergent iteration. Hence, we must also have that $S^{\text{FS}} \rightarrow S^n$. We therefore define an error in the source function by:

$$E^n = S^{\text{FS}} - S^n. \quad (41)$$

This error term is used for calculating approximate corrections to the radiation field, and these corrections approach zero when the error term approaches zero. We rewrite the error by using the inverted version of Eq. (39) for S^n and Eq. (40) for S^{FS} :

$$E^n = \rho \int_{\Omega} p_{\Omega}^{\text{FS}} \frac{d\Omega}{2\pi} + \eta - \Lambda_{\Omega}^{-1}[p_{\Omega}^{\text{FS}}]. \quad (42)$$

The exact correction $\delta p_{\Omega} = p_{\Omega} - p_{\Omega}^{\text{FS}}$ is added to p_{Ω}^{FS} in this equation, which by Eq. (36) results in $E^n = 0$. This can be written as:

$$\rho \int_{\Omega} (p_{\Omega}^{\text{FS}} + \delta p_{\Omega}) \frac{d\Omega}{2\pi} + \eta - \Lambda_{\Omega}^{-1}[p_{\Omega}^{\text{FS}} + \delta p_{\Omega}] = 0. \quad (43)$$

Subtraction of Eq. (42) from Eq. (43), and using the fact that Λ_{Ω}^{-1} is linear since we have assumed that the opacity χ is independent of the radiation field, yields an equation for the exact correction δp_{Ω} :

$$\Lambda_{\Omega}^{-1}[\delta p_{\Omega}] - \rho \int_{\Omega} \delta p_{\Omega} \frac{d\Omega}{2\pi} = E^n. \quad (44)$$

The angle integral results in a coupled set of second order differential equations as for the original problem in Eq. (36), and is equally difficult to solve. We therefore make a simplification in the angle integral, and the solution of the equation will now define approximate instead of exact corrections. This leads to an iteration series that we must provide convergence for. The accuracy of the simplified angle integral determines the convergence rate. The error term E^n includes S^{FS} which is calculated using the full angle integral in Eq. (40), such that S^n will converge to the correct solution.

We assume that the corrections δp_{ω} in an arbitrary direction ω in a subset of ray directions $\Delta\Omega_c$ are comparable to the angle averaged correction:

$$\delta p_{\omega} \sim \int_{\Omega} \delta p_{\Omega} \frac{d\Omega}{2\pi} = \delta J \quad ; \omega \in \Delta\Omega_c. \quad (45)$$

Since any direction $\omega \in \Delta\Omega_c$ gives approximately the same correction by this assumption, it is only necessary solve for the approximate correction without coupling to other rays in $\Delta\Omega_c$. The equation for the corrections is defined by inserting Eq. (45) in Eq. (44):

$$\Lambda_{\omega}^{-1}[\delta p_{\omega}^*] - \rho \delta p_{\omega}^* \equiv E^n. \quad (46)$$

This approach is exact is the special case of optically thick media, for which any direction gives the same perturbation due to isotropy. In optically thin three dimensional media, the solutions from Eq. (46) will be slightly different for different ω . We therefore solve the equation on several rays in $\Delta\Omega_c$, and the angle average of these perturbations give a more accurate estimate of the perturbation in mean intensity:

$$\delta J^* = \frac{1}{\Delta\Omega_c} \int_{\Delta\Omega_c} \delta p_{\omega}^* d\omega \simeq \int_{\Omega} \delta p_{\Omega} \frac{d\Omega}{2\pi} = \delta J. \quad (47)$$

For the special case of planetary or stellar atmospheres which have roughly a plane parallel character, we could use a one point angular quadrature in zenith angle (Ed-

dington approximation for plane parallel atmospheres), and choose a set of points in azimuth angle to define $\Delta\Omega_c$. This corresponds to a bidirectional cone symmetrical about the vertical line.

The approximate perturbation in mean intensity δJ^* is now used to find the next estimate of the source function S^{n+1} . Eq. (43) for the exact correction can be rewritten as:

$$S^{\text{FS}} + \rho \delta J - S = 0.$$

Hence, the next estimate S^{n+1} of S becomes:

$$S^{n+1} = S^{\text{FS}} + \rho \delta J^* = S^{\text{FS}} + \delta S^*. \quad (48)$$

The iteration steps can be summarized as follows (using the left hand side results on the right hand side in the following line):

$$\begin{aligned} S^{\text{FS}} &= \rho \Lambda[S^n] + \eta \\ E^n &= S^{\text{FS}} - S^n \\ \Lambda_{\omega}^{-1}[\delta p_{\omega}^*] - \rho \delta p_{\omega}^* &= E^n \\ \delta J^* &= \frac{1}{\Delta\Omega_c} \int_{\Delta\Omega_c} \delta p_{\omega}^* d\omega \\ S^{n+1} &= S^{\text{FS}} + \rho \delta J^*. \end{aligned}$$

3.3. Comparison to other 3D iterative methods

It is now instructive to discuss other methods on the same scattering problem and compare them to the present method. It will then become clear that the present method is favorable for the coherent isotropic scattering problem.

Cannon's AQPT method. The Angle Quadrature Perturbation Technique (AQPT) of Cannon (1973a) is related to the present method in that the angle integral is approximated, and that the integro-differential approach is used as a basis. The angle integral for the corrections is represented by a quadrature sum with fewer angles (lower order quadrature) than the more accurate quadrature sum used for the formal solution of the mean intensity needed in the error term. All rays in the lower order quadrature sum are used in the correction procedure, and this corresponds to a set of coupled differential equations. This is computationally demanding in 3D since it leads to a matrix equation with a complex matrix structure.

Cannon's method results if we instead of the one-ray approximation in Eq. (45) use an approximate angle integral with few quadrature points Ω_i . This replaces the angle integral¹ in Eq. (44):

$$\Lambda_{\Omega_i}^{-1}[\delta p_{\Omega_i}^*] - \rho \sum_i \delta p_{\Omega_i}^* w_i = E^n. \quad (49)$$

Here, w_i are the quadrature weights. All perturbations $\delta p_{\Omega_i}^*$ are found simultaneously, instead of finding them suc-

¹ The symbols of continuous integrals means that they are implemented numerically as high order quadrature sums. The approximate angle integral is written explicitly as a quadrature sum to distinguish it from the more accurate angle integral.

cessively, ray by ray. The perturbation in mean intensity is given by:

$$\delta J^* = \sum_i \delta p_{\Omega_i}^* w_i. \quad (50)$$

As a special case of Cannon's method, we can use only one ray Ω_0 in the approximate angle integral that gives the one-ray approximation that we have used. However, this corresponds to $\delta J^* = \delta p_{\Omega_0}^*$ by Eq. (50), and is not the same as the present method, since we average over several solutions of the correction equation (Eq. 46) to get the corrections in mean intensity.

Nordlund's method. Nordlund (1985) developed a method for spectral line radiation in CRD for two level atoms. Nordlund uses a similar one-ray approximation as in Eq. (45), and the correction in mean intensity is also in this case calculated after δp_{Ω}^* along individual rays has been solved for. Instead of the differential equation approach, he uses an integral equation approach. Here, we reformulate Nordlund's method for the present coherent problem. A ray dependent error is defined by:

$$E_{\Omega}^n = p_{\Omega}^{\text{FS}} - p_{\Omega}^n. \quad (51)$$

If the exact correction $\delta p_{\Omega} = p_{\Omega} - p_{\Omega}^n$ is added to p_{Ω}^n in this equation, we should get $E_{\Omega}^n = 0$. This can be written as:

$$\Lambda_{\Omega}[\rho \int_{\Omega} (\delta p_{\Omega} + p_{\Omega}^n) \frac{d\Omega}{2\pi} + \eta] - \delta p_{\Omega} - p_{\Omega}^n = 0. \quad (52)$$

Subtraction of Eq. (52) from Eq. (51), yield an equation for the exact correction δp_{Ω} :

$$\delta p_{\Omega} - \Lambda_{\Omega}[\rho \int_{\Omega} \delta p_{\Omega} \frac{d\Omega}{2\pi}] = E_{\Omega}^n.$$

Nordlund assumes isotropic perturbations:

$$\int_{\Omega} \delta p_{\Omega} \frac{d\Omega}{2\pi} \simeq \delta p_{\Omega},$$

and the resulting approximate correction equation is solved for ray by ray:

$$\delta p_{\Omega}^* - \Lambda_{\Omega}[\rho \delta p_{\Omega}^*] = E_{\Omega}^n. \quad (53)$$

The correction to the source function is:

$$\delta S^* = \rho \int_{\Omega} \delta p_{\Omega}^* \frac{d\Omega}{2\pi},$$

and the next estimate for the source function is:

$$S^{n+1} = \delta S^* + S^n.$$

Local ALI method. Eq. (37) for the volume integral form of the problem has been solved by iterations using the well known Accelerated Lambda Iteration technique (ALI) (e.g., Hubeny 1992). For 3D problems, it is most convenient to use the diagonal (local) elements Λ_{ii} of the kernel matrix as the approximate operator, since this involves only scalar operations for the corrections to the source function. This scheme is in our case identical to Jacobi iterations (Jacobi 1845) on a system of linear equations. In practice, one uses the approximate diagonal elements,

that are computed using an analytical approximation in combination with a numerical evaluation (e.g., Olson et al. 1986). The corrections to the source function in the local ALI scheme is in our notation:

$$\delta S^* = \frac{E^n}{1 - \rho \Lambda_{ii}^*},$$

where Λ_{ii}^* is the approximation of the diagonal elements. This operation is local, i.e. the corrections are calculated by multiplying the error term by a factor, grid point by grid point. The updated value is $S^{n+1} = \delta S^* + S^n$ as for the lambda operator method of Nordlund.

Discussion of methods. The ALI method is characterized by approximations in spatial coupling and accurate angle coupling, while Cannon's, Nordlund's and the present method are characterized by approximations in angle coupling and accurate spatial coupling.

The ALI method has much slower convergence rate than both Nordlund's and the present method, since it does not provide global coupling along rays. The necessary number of iterations to obtain the desired accuracy for ALI is large in comparison to the ray methods, and the time spent on formal solutions for the error term will therefore be correspondingly larger. Cannon's method produces too large matrices in 3D since it involves coupled differential equations through the approximate angle integral. We will therefore not discuss the ALI and Cannon methods in further detail, but concentrate on Nordlund's method and the new method.

Eq. (46) for the corrections can be written:

$$W[\delta p_{\omega}^*] \equiv (\Lambda_{\omega}^{-1} - \rho)[\delta p_{\omega}^*] = E^n. \quad (54)$$

Hence, W can be written as a matrix by subtracting ρ from the diagonal elements in the standard Feautrier matrix. The Feautrier matrix is tridiagonal, and this results from using finite differences for the second derivative in the Λ_{Ω}^{-1} operator (Eq. 32). The matrix form of W is therefore also tridiagonal and the solution of $W[\delta p_{\omega}^*] = E^n$ requires only on the order of D arithmetic operations, where D is the number of points along a ray. The motivation for developing the method was to preserve the tridiagonal structure to minimize computational effort.

The correction equation in Nordlund's method (Eq. 53) contains the Λ_{Ω} operator. The matrix form of Λ_{Ω} is filled with nonzero elements since the integrals of the lambda operator in Eq. (33) have in general geometrically extended kernels. The matrix for solving the corrections is therefore also filled, after the modifications indicated by Eq. (53). The Λ_{Ω} matrix contains D^2 elements where D is the number of samples along the ray. It would take D formal solutions along the ray to calculate it directly by shifting a delta function (for the source function) D times along the ray, and D^3 arithmetic operations to solve for the corrections.

A more efficient procedure (Nordlund 1985 and Nordlund 1991) uses Scharmer's approximate lambda opera-

tor $\Lambda_{\Omega}^{\dagger}$ (Scharmer 1981) which is based on the one point Gauss-Laguerre quadrature for each of the integrals in Eq. (33). This results in a sparse matrix with two bands around the diagonal for the coherent problem. The number of arithmetic operations for solving the correction equation with Scharmer's operator is between D^2 and D^3 , since advantage can be taken of the sparse matrix.

This means that the new tridiagonal-based method saves at least a *factor* D (typically 100) in arithmetic operations for the corrections. In addition, we do not need to calculate approximate operators, since the differential operator W is exact. It must be noted that the 2-level line problem with frequency redistribution (Nordlund 1985 and Nordlund 1991) is more efficiently solved with Scharmer's operator, than using the differential operator approach. But for the coherent problem, we obtain a considerable saving of computer time with the new method.

4. Application to a solar simulation

We have extended the solar convection model of Nordlund & Stein (1990) vertically to include chromospheric layers. The previous version of the code included the convective overshoot zone in the photosphere, and the upper boundary was located at 500 km above the surface (cooling layer) of the convection zone. Radiation was treated in LTE by using averaged Planck functions in a few wavelength groups as source functions, according to the procedure explained in Nordlund (1982).

The new model extends to 1500 km above the surface. This gives a coverage of the photosphere and a large part of the chromosphere. The horizontal extension is 6×6 Mm, typically covering five large granular cells. The lower boundary is located 1500 km below the top of the convection zone. This gives a total vertical extension of 3 Mm, covering two orders of magnitude in density in the convection zone, and five orders of magnitude in the atmosphere. This model enable us to study the dynamic coupling between the convection zone and the atmosphere above. Hydrodynamic results from these simulations are treated in a separate paper (Skartlien et al. 1998).

We solve the radiation scattering problem from the top of the domain down to a layer within the convective part of the simulation. This zone covers about ten orders of magnitude in opacity. For each timestep and wavelength group, we iterate the solution to convergence, and start the iteration for the next timestep with the solution from the previous timestep. After convergence of all groups, we calculate the radiative flux divergence for the gas-energy equation.

In this section, we explain the use of lookup tables, and how line scattering opacity is treated and how line opacity in the ultraviolet is included. We also give the radiation boundary conditions for this specific problem, and details in the implementation of the numerical method.

4.1. Calculation of group mean opacities

Lookup table. The scattering albedo, thermal source, and opacity per wavelength group are pre-calculated, and stored in a lookup table. These quantities are read from the table at each timestep of the simulation. We include these new quantities in the old lookup table used in Nordlund (1982).

Since the opacity is assumed to be in LTE, each radiation quantity is a function of two thermodynamic variables. The lookup table consists, per wavelength group i , of the quantities $\rho_i(e, \ln(\rho))$ and $\eta_i(e, \ln(\rho))$ for the source function in Eq. (22), and $\chi_i^T(e, \ln(\rho))$ in Eq. (23) for the transport equation. All three quantities are a function of specific internal energy e [erg g⁻¹ cm⁻³] and the natural logarithm of density ρ [g cm⁻³]. These two variables are taken from solutions of the energy equation and the mass conservation equation. For each spatial point in the simulation, we get a set of radiation quantities by looking up the values from the table.

For group mean opacities that are calculated by averaging over an approximate radiation field, we use the one dimensional mean intensity J_{λ}^{PP} of the horizontally averaged atmosphere. J_{λ}^{PP} is only a function of the height z or alternatively density $\bar{\rho}(z)$ in the averaged atmosphere. We use the one dimensional relation $J_{\lambda}^{\text{PP}}(\bar{\rho})$ to find the radiation field to be used in the table by assuming $J_{\lambda}^{\text{PP}}(\rho)$ where ρ is now the density from the 3D simulation. Fluctuations in ρ at fixed height corresponds to a height error in the sampling of J_{λ}^{PP} of about ± 100 km.

The weighting between streaming and diffusion regimes in Eq. (25) is done by setting the typical length scale of variation of the group mean Planck function l_i equal to the typical density scale height in the photosphere, which is 150 km.

Monochromatic continuum and line opacities. Continuum opacities are calculated using the package of Gustafsson (1973). Absorption opacities κ_{λ}^c are calculated for bound-free and free-free transitions of H⁻ and H I, for bound-free transitions of H₂⁻, H₂, H₂⁺, He I, C I, Mg I, Al I and Si I. Scattering opacities σ_{λ}^c are calculated for Rayleigh and electron scattering. Electrons come from ionized stages of the most abundant elements H, He, C, N, O, Ne, Na, Mg, Al, Si, S, K, Ca, Cr, Fe and Ni.

Line opacities χ_{λ}^1 (sum of absorption opacities κ_{λ}^1 and scattering opacities σ_{λ}^1) for CO and CN molecules, and for the strongest lines of the most abundant elements are taken from opacity distribution functions by Gustafsson et al. (1975). The opacity spectrum consists of 368 wavelength points from 153.3 nm in the ultraviolet to 12.43 μm in the infrared.

Line opacity in ultraviolet. Line opacities χ_{λ}^1 from Fe in ultraviolet are accounted for by refining the empirical fit

of Magain (1983). The ratio between the approximate UV-opacity and H^- continuum opacity can be written as:

$$\begin{aligned} \frac{\chi_\lambda^1(\text{UV})}{\chi_\lambda^c(\text{H}^-)} &= \alpha \frac{k_\lambda}{A_\odot} \frac{n_i}{n_{H^-}} \\ &= \alpha \frac{k_\lambda}{A_\odot} \frac{n_i}{n_{\text{Fe I}}} \frac{n_{\text{Fe I}}}{n_{\text{Fe II}}} \frac{n_{\text{Fe II}}}{n_{\text{Fe}}} A_{\text{Fe}} n_{\text{H}} \Big/ \frac{n_{\text{H}^-}}{n_{\text{H}}} n_{\text{H}}. \end{aligned} \quad (55)$$

The constant α depends upon partition functions, statistical weights and physical constants in the Saha-Boltzmann equations. A_\odot and A_{Fe} are respectively the solar abundance and a general abundance of Fe relative to H. All n are number densities for the species indicated, and n_i is the population density of a mean Fe I energy level chosen to have 3 eV excitation potential (Holweger 1970). k_λ is the empirical function to be determined. Magain assumes that almost all Fe is singly ionized:

$$n_{\text{Fe I}} \ll n_{\text{Fe II}} \simeq n_{\text{Fe}},$$

and derives a relation for k_λ by setting $n_{\text{Fe II}}/n_{\text{Fe}} = 1$ in Eq. (55). k_λ is then determined by forcing the flux from a solar model to fit observations. We use this empirical k_λ . However, the ionization assumption leads to a strong overestimate of the UV-opacity at low temperatures for which $n_{\text{Fe II}}$ is small. To prevent this overestimation is cool regions in the chromosphere ($T \sim 2000$ K), we use:

$$\frac{n_{\text{Fe I}}}{n_{\text{Fe II}}} \frac{n_{\text{Fe II}}}{n_{\text{Fe}}} \simeq \frac{n_{\text{Fe I}}}{n_{\text{Fe I}} + n_{\text{Fe II}}}.$$

We calculate $\chi_\lambda^c(\text{H}^-)$ and determine $\chi_\lambda^1(\text{UV})$ from Eq. (55) when constructing the lookup table.

Line destruction probability. The contributions to line opacity from absorption is calculated using the approximation of Van Regemorter (1962). The destruction probability in lines ϵ_λ^1 is in general dependent upon several electronic transitions k on the same wavelength with individual destruction probabilities $\epsilon_\lambda^{1,k}$:

$$\epsilon_\lambda^1 = \frac{\sum_k \chi_\lambda^{1,k} \epsilon_\lambda^{1,k}}{\chi_\lambda^1} = \frac{\kappa_\lambda^1}{\chi_\lambda^1}, \quad (56)$$

where:

$$\chi_\lambda^1 = \sum_k \chi_\lambda^{1,k}$$

is the total line opacity, and $\chi_\lambda^{1,k}$ the line opacity at a transition k . Hence, this weighting procedure favors those transitions with the largest probability to interact with photons. The destruction probability in transition k is given by the probability for collisional de-excitation from the upper level j to the lower level i :

$$\epsilon_\lambda^{1,k} = \frac{1}{\frac{A_{ji}}{C_{ji}} + 1}.$$

A_{ji} [s^{-1}] is the Einstein coefficient for spontaneous radiative de-excitation, and C_{ji} [s^{-1}] the collisional de-excitation parameter.

We use Van Regemorter's approximation for C_{ji}/A_{ji} that is independent of the specific transition:

$$\frac{C_{ji}}{A_{ji}} = 20.6 \lambda^3 n_e T^{-1/2} P_{N,I} \left(\frac{\Delta E}{kT} \right), \quad (57)$$

where λ [cm] is the photon wavelength, n_e [cm^{-3}] electron number density, T [K] temperature, ΔE [erg] the photon energy, k [erg K^{-1}] the Boltzmann constant. The function P is given either for positive ions or neutrals. We use a monotonic spline (Steffen 1990) to interpolate between tabulated values for P . Using this approximation, we can write ϵ_λ^1 as two sums, one for neutrals N and one for ions I :

$$\epsilon_\lambda^1(n_e, T) = \frac{\epsilon_\lambda(N) \sum_k \chi_\lambda^{1,k}(N) + \epsilon_\lambda(I) \sum_k \chi_\lambda^{1,k}(I)}{\chi_\lambda^1},$$

or in terms of the opacity fraction $X_\lambda(n_e, T)$ from neutrals:

$$\epsilon_\lambda^1(n_e, T) = \epsilon_\lambda(N) X_\lambda + \epsilon_\lambda(I) \{1 - X_\lambda\}.$$

In the test examples in Sect. 5 and Sect. 6, we set the neutral fraction to unity as a first approximation. This gives an underestimation of the destruction probability, since the P-function has smaller magnitude for neutrals than for ions (cf. Van Regemorter 1962, Fig. 4). This enhances the line scattering contribution, compared to a weighting between the different P-functions by the opacity fractions. Van Regemorter's approximation is only useful for order of magnitude estimates (e.g., Mason & Monsignori-Fossi 1994), and it is doubtful if more realistic values for the destruction probability is obtained by the weighting given in the last equation.

We use the estimate for line destruction probability $\epsilon_\lambda^1(n_e, T)$ to calculate line absorption opacities κ_λ^1 from Eq. (56). Line opacities χ_λ^1 are taken from opacity distribution functions and Magain's approximation in ultraviolet.

4.2. Calculation of radiative flux divergence

Boundary conditions. It is a natural choice to use periodic horizontal boundaries in models of stellar or planetary atmospheres, as long as the simulation domain includes the largest structures that can be expected. Imposed boundary values are then not necessary at the sidewalls of the domain, but we must treat outgoing intensity at one sidewall as incoming intensity at the opposite sidewall. Specification of intensities are necessary only at the upper and lower boundaries.

The incoming intensity I_1^- at the upper boundary is estimated by assuming constant source function above this boundary:

$$I_1^- = \int_0^{\tau_1} S e^{-t} dt \simeq S_1(1 - e^{-\tau_1}).$$

The incoming intensity I_D^+ at the lower boundary is estimated by using the diffusion approximation as in Eq. (16):

$$I_D^+ = S_D + \frac{\partial S_D}{\partial \tau_\Omega} \simeq S_D + \frac{S_D - S_{D-1}}{\Delta \tau_{\Omega, D-\frac{1}{2}}}.$$

Discretization. The extent of the computational domain determines the coordinate system to be chosen due to spherical or elliptical geometry in stellar or planetary atmospheres. Here we have used Cartesian coordinates, which is suitable for a sufficiently small domain where the curvature can be neglected. We choose equal spacing between gridpoints in the horizontal directions. Uneven spacing is used vertically to resolve gradients in the photosphere.

We use the long characteristics method for radiation, with sampling only on horizontal planes (Nordlund 1985), such that every ray consists of the N_z depth points when using periodic horizontal boundaries. This results in relatively simple algorithms suitable for parallel processing since there are no computational differences for varying ray direction. The short characteristics (SC) method (Kunasz & Auer 1988) has better resolution along slanted rays, since also vertical planes are sampled. We cannot use the SC-method within the present ray-based correction procedure.

We use standard second order centered differences in the Λ_Ω^{-1} -operator, and in Eq. (46) for the corrections. For the high resolution angle integral over p_Ω^{FS} (for S^{FS} in the error term E^n), we use Gaussian quadrature integration over zenith angle, and trapezoidal integration with even spacing over azimuthal angle. This choice corresponds to rays along a set of bi-directional cones centered about the vertical line. We follow Nordlund (1982), and rotate this bundle of rays in azimuthal direction with a small increment each timestep. This reduces systematical errors due to angle discretization.

All rays in the cone with smallest zenith angle defines our subset of angles in which we solve the correction equation. This choice gives faster convergence rate than using rays with larger angles from the vertical line.

Numerical conditioning for radiation. For the formal solution in optically thin layers, we solve the Feautrier equation:

$$\Lambda_\Omega^{-1}[p_\Omega^{\text{FS}}] = S^n. \quad (58)$$

We solve for the difference $q_\Omega^{\text{FS}} = p_\Omega^{\text{FS}} - S^n$ in optically thick layers where $p_\Omega^{\text{FS}} \simeq S^n$. The differential equation for q^{FS} is obtained by taking the second derivative in optical path length τ_Ω of Eq. (58):

$$\Lambda_\Omega^{-1}[q_\Omega^{\text{FS}}] = \frac{\partial^2 S^n}{\partial \tau_\Omega^2}. \quad (59)$$

A discussion of the numerical advantages by this approach can be found in Nordlund & Stein (1990).

The flux divergence integral (per group) is calculated by quadrature integration with full angle resolution. We have, according to Eq. (27):

$$\nabla \cdot \mathbf{F}_i^* = \sum_{\Omega} \frac{\Delta \tau_\Omega}{\Delta r} (S - p_\Omega) w_\Omega. \quad (60)$$

Here, w_Ω are the angle quadrature weights, p_Ω and S are the converged solutions, $\Delta \tau_\Omega$ is the optical path length, and Δr the distance along a ray between two adjacent horizontal planes. To provide consistency with the Feautrier scheme, we use the ratio $\Delta \tau_\Omega / \Delta r$ for the mean opacity along the ray segment.

Interpolation between rays and Cartesian grid. We have $N_x \times N_y$ parallel rays passing through each horizontal plane for each ray direction. (N_x and N_y are the number of gridpoints in the horizontal directions). These rays do not necessarily pass through grid points, and we must therefore interpolate values from grid points onto rays. We also need to do the reverse interpolation from rays to grid. We use a local spline interpolation procedure as in Nordlund (1982).

These interpolations are not reversible, as opposed to the Fourier interpolations used in Nordlund (1985) and Kislman & Nordlund (1995). The aliased components in the discrete Fourier transform are sufficient to introduce an instability in the iteration procedure. This could be the reason why Kislman & Nordlund (1995) applied a stabilizing ‘‘extra’’ lambda iteration in their scheme. The high spatial frequencies are successively amplified if not a lowpass (de-aliasing) filter is applied.

We have therefore used the more robust spline interpolation. This need not to be reversible as long as we avoid to interpolate the same quantity back and forth between grid and rays during the iteration. This situation arises in the deepest layers where the specific intensity is equal to the source function along the rays. In this way, the next estimate of the source function should be equal to the previous estimate and the error term would be zero. However, this is not true for non-reversible interpolations. We solve this problem by interpolating q_Ω^{FS} back from the rays in the deepest layers. Since this is zero in these layers we ensure an error term that is zero.

Let a set of parallel rays pass through the grid points at a chosen reference plane z_0 . Let a ray coincide with the Cartesian gridpoint (x_0, y_0) in this reference plane. Let $\mathbf{r} = [r_x, r_y]$ be the displacement vector in another plane z , from (x_0, y_0) out to a point (x, y) where the ray intersects the plane. We have:

$$r_x = \Delta z \frac{dx}{dz},$$

and

$$r_y = \Delta z \frac{dy}{dz}.$$

$\Delta z = z - z_0$ is the vertical distance from the plane z_0 , to the plane z where we want to interpolate. The derivatives are given by:

$$\frac{dx}{dz} = \tan(\theta) \cos(\phi),$$

and

$$\frac{dy}{dz} = \tan(\theta) \sin(\phi),$$

where θ is the angle from the z -axis, and ϕ the angle from the x -axis. Hence, the displacement vector is a function of ray direction (θ, ϕ) and the vertical position of the plane: $r = r(\theta, \phi, z)$.

With given r , we use a successive local cubic spline interpolation based on four points. First we interpolate for the displacement r_x , then we use these values to calculate the interpolation corresponding to the displacement r_y . For either direction, let $p = (r - R)/\Delta$, where r is one of the displacement components r_x or r_y , and R is the nearest integer Cartesian coordinate less than r in direction x or y . Δ is the grid spacing in either direction. The cubic spline coefficients are:

$$\begin{aligned} q &= 1 - p \\ b_1 &= q + pq(q - p) \\ b_2 &= p - pq(q - p) \\ a_1 &= -\frac{1}{2}pq^2 \\ a_4 &= -\frac{1}{2}qp^2 \\ a_2 &= b_1 - a_4 \\ a_3 &= b_2 - a_1. \end{aligned}$$

The interpolation formula for one direction is:

$$f = a_1 g_{k-1} + a_2 g_k + a_3 g_{k+1} + a_4 g_{k+2},$$

where f is the interpolant to be evaluated between grid-lines k and $k + 1$, and g is the function to be interpolated. Since the grid spacing is uniform in the horizontal directions, we have only one set of spline coefficients for the same plane. The same procedure is carried out when we interpolate from rays to grid points, but now with a sign reversal of the displacement vector.

Implementation of iterations. Let “g” be grid variables and “r” ray variables, and let arrows denote interpolations between grid and rays. We suppress the angular index Ω of p^{FS} and q^{FS} for clarity. We use logarithmic interpolation for quantities that are always positive (such as intensity) to avoid unphysical negative interpolants.

For given S_g^n , we calculate the error E_g^n by the following procedure:

$$\log(\chi_g) \rightarrow \log(\chi_r)$$

$$\log(S_g^n) \rightarrow \log(S_r^n).$$

After the solution of Eq. (58) for optically thin layers:

$$\log(p_g^{\text{FS}}) \leftarrow \log(p_r^{\text{FS}}).$$

After the solution of Eq. (59) in optically thick layers:

$$q_g^{\text{FS}} \leftarrow q_r^{\text{FS}}.$$

We calculate the formal solution S_g^{FS} by running through all ray directions by repeated usage of the last four equations. For optically thick layers:

$$S_g^{\text{FS}} = \rho_g \left(\sum_{\Omega} q_c^{\text{FS}} w_{\Omega} + S_g^n \right) + \eta_g.$$

For optically thin layers:

$$S_g^{\text{FS}} = \rho_g \sum_{\Omega} p_c^{\text{FS}} w_{\Omega} + \eta_g.$$

The error is now calculated (Eq. 41) as:

$$E_g^n = S_g^{\text{FS}} - S_g^n.$$

We can now calculate the correction to the source function δS_g^* . These interpolations are done for the corrector step:

$$\log(\chi_g) \rightarrow \log(\chi_r)$$

$$E_g^n \rightarrow E_r^n.$$

We have that $\epsilon_g = 1 - \rho_g$, where ρ_g is the scattering albedo and ϵ_g the destruction probability. Since $\epsilon_g \in \langle 0, 1 \rangle$, we must ensure that the same is true for the interpolant. To do this, we use these transformations:

$$a_g = \frac{\epsilon_g}{1 - \epsilon_g}$$

$$\log(a_g) \rightarrow \log(a_r)$$

$$\epsilon_r = \frac{a_r}{1 + a_r} \in \langle 0, 1 \rangle.$$

The corrections δp_r^* are now calculated from Eq. (46). Then we interpolate these corrections back to the grid:

$$\delta p_g^* \leftarrow \delta p_r^*.$$

δS_g^* is calculated by using the last six equations while running through the subset of angles. The quadrature sum over the subset of angles gives:

$$\delta S_g^* = \rho_g \sum_{\omega} \delta p_g^* w_{\omega}.$$

w_{ω} are the angle quadrature weights. The updated source function is by Eq. (48):

$$S_g^{n+1} = S_g^{\text{FS}} + \delta S_g^*.$$

We can now go back to the calculation of the error term for the next iteration.

After convergence of this scheme, we calculate the flux divergence contributions along each ray and interpolate these back to the grid:

$$\left\{ \frac{\Delta \tau_{\Omega}}{\Delta r} (S - p_{\Omega}) \right\}_g \leftarrow \left\{ \frac{\Delta \tau_{\Omega}}{\Delta r} (S - p_{\Omega}) \right\}_r.$$

The flux divergence per group is obtained by the angle integration in Eq. (60). The total flux divergence is obtained by adding over all groups according to Eq. (28):

$$\nabla \cdot \mathbf{F}^* = \sum_i \sum_{\Omega} \left\{ \frac{\Delta \tau_{\Omega}}{\Delta r} (S - p_{\Omega}) \right\}_g w_{\Omega}.$$

CPU time as function of resolution. The CPU time spent on spline interpolations per ray direction scales linearly with $N_{\text{grid}} = N_x N_y N_z$. This is also the case for the formal solutions p_{Ω}^{FS} along rays, since there are $N_x N_y$ rays per direction, and the solution of p_{Ω}^{FS} per ray scales as N_z . The CPU time for the formal solution S^{FS} is therefore:

$$T_{\text{FS}} = a N_{\text{rays}} N_{\text{grid}},$$

where a is a constant. N_{rays} is the number of ray directions. Here, we have not taken into account external arithmetical operations since they consume a negligible amount of time in comparison. Similarly, the time spent on corrections δS^* is:

$$T_{\text{corr}} = b(\alpha N_{\text{rays}})N_{\text{grid}},$$

where $b \simeq a$ is a constant, and $\alpha = N_{\text{subset}}/N_{\text{rays}} < 1$, where N_{subset} is the number of rays in the correction procedure. $\alpha = 1/3$ when using one set of azimuthal angles from 3 sets defined by a 3 point Gaussian quadrature in zenith angle. The CPU-time for one iteration step in one wavelength group is $T_{\text{it}} = T_{\text{FS}} + T_{\text{corr}}$. Hence, the total CPU-time for solving for the radiation also scales linearly with the number of ray directions and gridpoints: $T_{\text{tot}} = T_{\text{it}}N_{\text{iter}} \sim N_{\text{rays}}N_{\text{grid}}N_{\text{iter}}$, where N_{iter} is the total number of iterations needed for all groups. Since $b \simeq a$ and $\alpha < 1$, most of the CPU time is used on formal solutions.

Parallel processing. We have used MPI (Message Passing Interface) on a CRAY ORIGIN 2000 machine, to compute formal solutions p_{Ω}^{FS} and corrections δp_{Ω}^* for different angles in parallel (for angle-parallelization in combination with the SC method, see Botnen 1997). For each iteration, the data needed for different angles is distributed to the set of microprocessors we use. The resulting ray data is thereafter gathered and averaged (angle integrals) in one microprocessor. This scheme is repeated until convergence of the iteration.

In principle, the *computational time* (not CPU time) will be reduced by a factor $1/N_{\text{rays}}$ and $1/N_{\text{subset}}$ for formal solutions and corrections respectively, if we use N_{rays} and N_{subset} microprocessors for formal solutions and corrections. In practice, the time reduction is less due to transfer of data between microprocessors; more microprocessors means more time for exchanging data. In some cases we therefore calculate, in succession, results for more than one ray direction per microprocessor. In these cases, the number of microprocessors is less than N_{rays} and N_{subset} . A further speed-up would result if we parallelized over groups in addition to ray directions.

5. Test of the methods

5.1. Monochromatic versus group mean solutions

We compare group mean solutions of radiative heating ϕ_i^* (negative of flux divergence), and mean intensity J_i^* with the corresponding exact solutions ϕ_i and J_i obtained from wavelength integration of monochromatic solutions. The monochromatic solutions are obtained from solving a 3D scattering problem at each of the 368 wavelength points we use. For this test, we have used a single snapshot from the three dimensional atmosphere-convection simulation, with the current radiation method included in the energy equation.

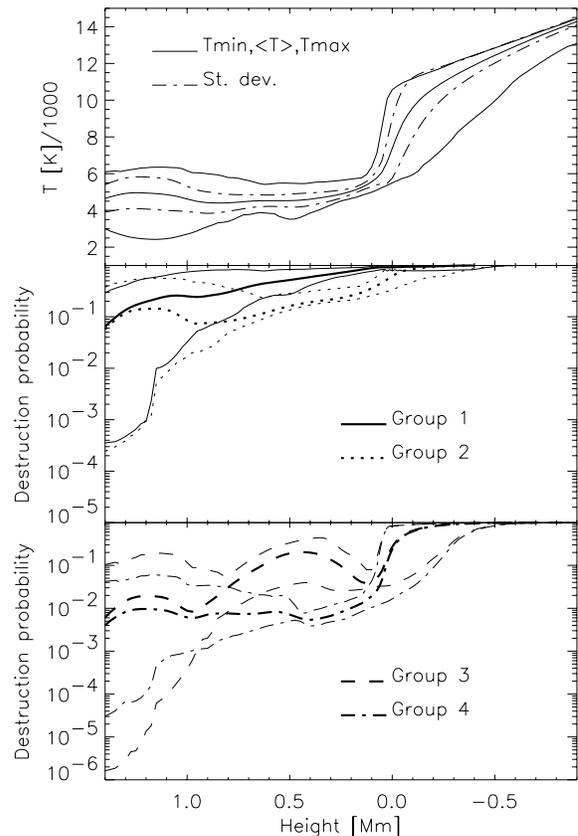


Fig. 1. *Upper panel:* Mean, maximum and minimum temperature (full drawn lines), and standard deviation (dash-dotted) in horizontal planes for the test-snapshot. *Lower panels:* Mean (thick lines), maximum and minimum (thin lines) group mean destruction probabilities in horizontal planes. As the line contribution increases (increasing group number), scattering becomes more important, and photons scatter rather than couple to the gas by collisions. The cooling layer of the convection zone is located around the height of 0.0 Mm, where the radiative cooling is largest.

We use four groups, denoted by 1, 2, 3 and 4, with group 1 as the continuum group, and group 4 for the strongest spectral lines. Intermediate groups correspond to weaker lines. Optical depths unity, corresponding to the group mean opacities, are located at increasing height for increasing group number due to larger average opacity for the spectral line groups. Group 1 transports most radiative flux, with maximum flux divergence in the photosphere or cooling layer of the convection zone. This group includes the largest number of wavelength points.

The temperature structure of the atmosphere and the group mean photon destruction probabilities are displayed in Fig. 1. The cooling layer of the convection zone is located around the height of 0.0 Mm, where the radiative cooling and mean temperature gradient are largest. The group mean destruction probability is given by $1 - \rho_i$,

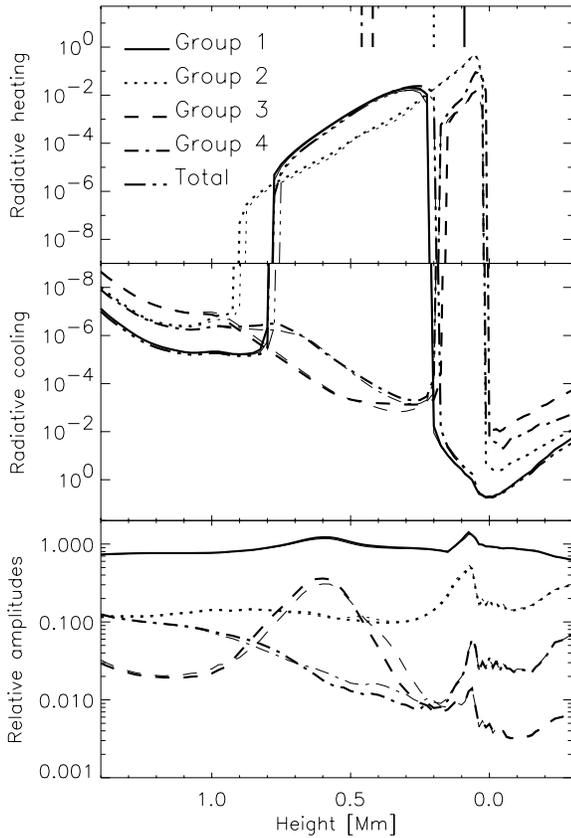


Fig. 2. *Upper panels:* Horizontally averaged radiative heating per volume unit shown in a bi-logarithmic plot. The units are arbitrary. Thick lines are the exact group solutions, while the thin lines are the approximate group solutions. The vertical line-segments in the uppermost panel show the heights where the average optical depth per group is unity, as measured along the vertical line. *Lower panel:* Horizontally averaged amplitudes of radiative heating (average absolute value) relative to the amplitude of the exact total heating. Thick lines are the exact amplitude ratios, while the thin lines are ratios from the approximate group solutions.

where the group mean scattering albedo ρ_i is given in Eq. (22).

As the line contribution to the group mean opacity increases (increasing group number), scattering becomes more important, and photons scatter rather than being coupled to the gas via collisions. For a given difference between local temperature and radiation temperature, smaller destruction probability means smaller amplitude of the radiative flux divergence as compared to no scattering. The scattering contribution is present also in the first few hundred kilometers into the convection zone (below 0.0 Mm), but it dominates in atmospheric layers above 0.0 Mm for groups 2, 3 and 4.

Horizontal averages of exact and approximate radiative heating for the four groups are compared in

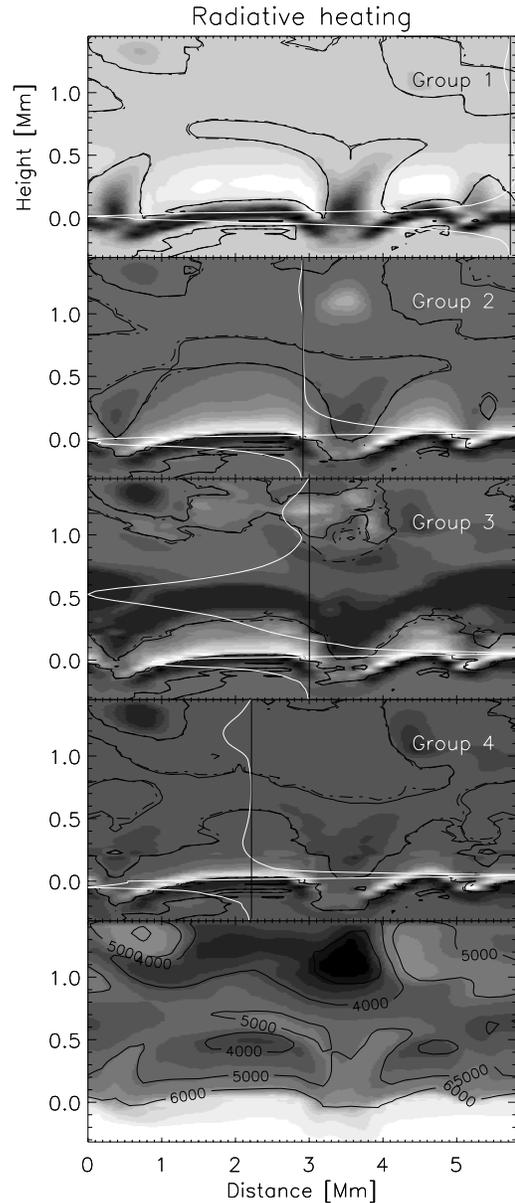


Fig. 3. Exact group radiative heating per mass unit in vertical slices through the simulation. Full drawn black contours show the level of zero heating. Heating is found at lighter shades of grey, and cooling at darker shades. Dash-dotted black contours show the zero level of the approximate group heating, and coincides well with the zero level of the exact heating. White curves are the horizontal averages of the radiative heating per mass unit (normalized to fit the plotting window). Black vertical lines marks the zero level for these curves, and positive values are to the right of the lines. *Lower panel:* Temperature in the same vertical slice. Note the cooling in the granular layer at the height 0.0 Mm, and heating immediately above in all groups, and also heating of the cool region below 3000 K in the chromosphere.

Fig. 2. Here we have used the full approximation given

by Eq. (22), Eq. (23) and Eq. (24), in the calculation of the approximate solution. The horizontal averages (thin curves) coincide very well with the horizontal averages of the exact solutions (thick curves). The upper panels show cooling below 0.0 Mm in all groups, and heating in groups 2, 3 and 4 immediately above 0.0 Mm.

The lower panel shows the amplitude of the approximate and exact flux divergence (horizontal average of the absolute value) relative to the amplitude of the *total exact* flux divergence, i.e., $\langle |\phi_i^*| \rangle / \langle |\phi_{\text{tot}}| \rangle$ and $\langle |\phi_i| \rangle / \langle |\phi_{\text{tot}}| \rangle$ (brackets denote horizontal average). We see that the exact relative amplitudes (thick curves) coincides very well with the approximate values (thin curves).

A sample of the spatial structure is displayed in Fig. 3, where we have shown the exact heating per mass unit in all groups as grey scale images in a vertical slice. Black contours mark locations of zero heating, and lighter shades of grey means positive heating. Gas in layers immediately above the cooling layer is heated in all groups in expanding flow above granules. Granules are seen as curved horizontal structures. As up-flowing gas expands and cools, the temperature falls below the radiation temperature, and the gas is radiatively heated. Note also heating of the cool region below 3000 K in the chromosphere, as radiation energy from below are converted to thermal energy via collisions.

The dash-dotted black contours show the zero level for the approximate heating, and coincides well with the zero level for the exact heating. White curves are horizontal averages per mass unit (normalized to fit the plotting window). The grey scale plots show that the heating (cooling) contributions can cancel a large part of the cooling (heating) contributions in the horizontal average at a specific depth. This is clearly the case for the curved granular surfaces.

The above comparisons of horizontally averaged heating do not reveal local deviations between approximate and exact solutions since e.g., excessive heating at a specific height and horizontal region can be cancelled by excessive cooling in other regions at the same height. We have therefore computed absolute differences between approximate and exact solutions at each gridpoint and thereafter calculated the horizontal average of these differences. The relative deviation is found by dividing by the magnitude of the exact radiative heating ϕ_i at the same height: $\langle |\phi_i^* - \phi_i| \rangle / \langle |\phi_i| \rangle$.

The third panel in Fig. 4 shows typically 1–30 % deviations in atmospheric layers, with a peak of 100 % deviation in group 3 at 0.35 Mm. The smallest deviations, less than 10 %, are found for group 1 that carries most energy. The total radiative heating is also seen to deviate with less than 10 % from exact value, due to the dominance of group 1 at all heights. This is the flux-divergence accuracy we can achieve in 3D simulations with the current method.

The fourth panel in Fig. 4 shows the fraction of gridpoints that have higher radiative heating than the exact

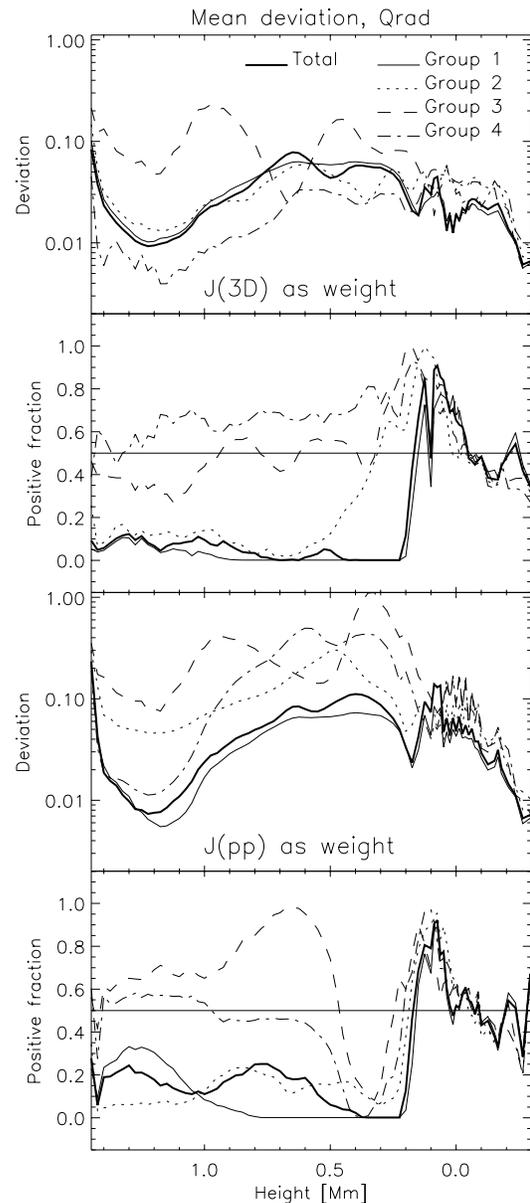


Fig. 4. Horizontally averaged deviations between approximate and exact group radiative heating. *First panel:* Group and total heating using the three dimensional monochromatic mean intensity as an opacity weighting function in the streaming regime. The deviations are typically in the range 1–15 %. *Second panel:* The fraction of gridpoints in horizontal planes for which the approximate solution is larger than the exact solution. *Third panel:* The plane parallel mean intensity from the horizontally averaged model is used as a weighting function. The deviations are somewhat larger when this weighting is used, typically in the range 1–30 %. *Fourth panel:* Same as for the second panel, but for the solutions using the plane parallel mean intensity as weighting function.

solution. A fraction of 0.5 corresponds to an approximate solution that fluctuates on either side of the exact so-

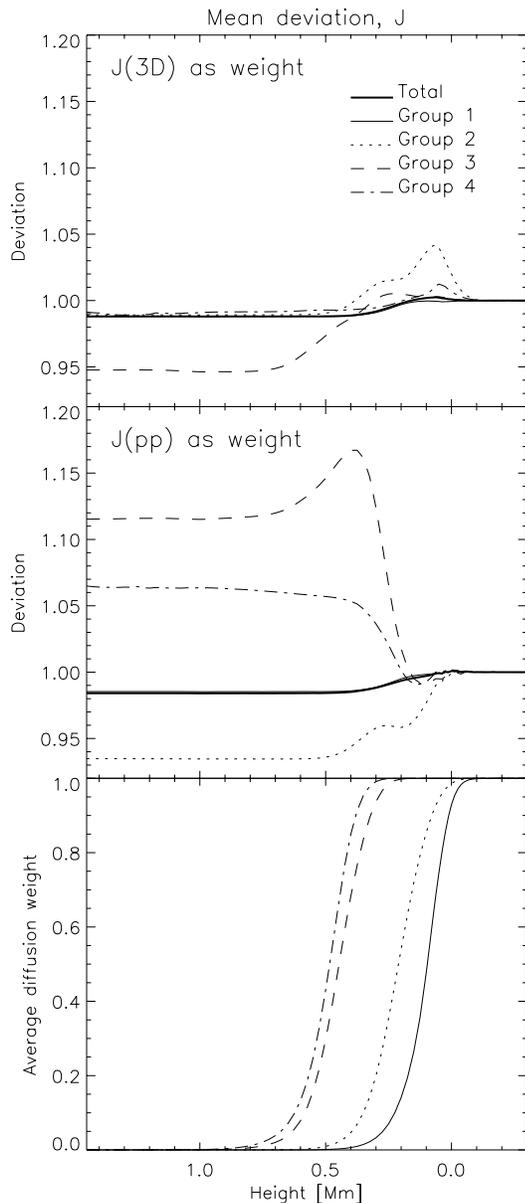


Fig. 5. Horizontally averaged ratios between approximate and exact group intensities. *Upper panel:* Group and total intensity using the three dimensional monochromatic mean intensity as an opacity weighting function in the streaming regime. The approximate solutions are on the average lower than the exact solutions in the streaming regime (above 0.5 Mm), as expected when using the mean intensity as a weight function rather than the specific intensity. *Middle panel:* The plane parallel mean intensity from the horizontally averaged model is used as a weighting function. *Lower panel:* Horizontally averaged diffusion weight. When this is unity, the group mean Rosseland opacity is used in both the upper and middle panel, and the deviations are negligible.

lution. We see that the approximate solutions has lower values than the exact solutions at 0.3 Mm (immediately

above the cooling layer). In the chromosphere, we have that group 1 and 2 have lower values, and group 3 higher values than the exact solution. Group 4 has approximately the correct values. The total radiative heating is lower than the exact value, due to the domination of group 1. By considering the deviation above, we conclude that the total heating is less than 10 % below the exact heating in atmospheric layers.

J_{λ}^{PP} versus J_{λ} as weight function. The plane parallel mean intensity from the horizontally averaged atmosphere, J_{λ}^{PP} , was used as an opacity weight function in the approximate solution discussed above. We note that this approach is the one we use in simulations. Here, we compare the exact solution with the approximate solution obtained by using the exact three dimensional mean intensity J_{λ} as a weighting function. This refinement does not lead to the exact group solutions, since this requires I_{λ} as weighting function. We modify Eq. (23) for the mean opacity in the lambda operator:

$$\chi_i^{\text{J}} \rightarrow \chi_i^{\text{R}} \equiv \chi_i^{\text{T}}.$$

The weighting between diffusion and streaming regimes is here the same as in Eq. (23). The opacity coefficients in the source function needs no weighting in this case since the three dimensional mean intensity naturally approaches the Planck function at depth. The source function is now a modified version of Eq. (22):

$$S_i^* = \left(\frac{\sigma_i^{\text{J}} \rightarrow \sigma_i^{\text{B}}}{\chi_i^{\text{J}} \rightarrow \chi_i^{\text{B}}} \right) J_i^* + \frac{\kappa_i^{\text{B}} B_i}{\chi_i^{\text{J}} \rightarrow \chi_i^{\text{B}}},$$

where the arrows are the natural transitions instead of the imposed weighting between diffusion and streaming regimes.

We note that the approximation in this source function enters in the streaming regime, since we use χ_i^{J} in the denominator rather than χ_i^{J} . This effect is however cancelled for the following reason. Using χ_i^{J} in the lambda operator, $J_i^* = \Lambda(\chi_i^{\text{J}})[S_i^*]$ cancels the same opacity in the source function when we consider the emissivity $\chi_i^{\text{J}} S_i^*$ (cf. the kernel function in Eq. (5) and the volume integral for the lambda operator). This means that we obtain correct thermal emissivity $\kappa_i^{\text{B}} B_i$, and scattering opacity in the emission term $\sigma_i^{\text{J}} J_i^*$. This test will therefore isolate the influence on radiation transport from using the approximate opacity χ_i^{J} (rather than the correct χ_i^{J}) that determines the optical path-lengths and exponential damping in the kernel function for the lambda operator.

The upper panel in Fig. 5 shows horizontally averaged ratios between approximate and exact mean intensities: $\langle J_i^*/J_i \rangle$. The approximate solutions are on the average lower than the exact solutions in the streaming regime (above 0.5 Mm), consistent with the arguments in Sect. 3 where we expected lower mean intensities resulting from $J_i^* = \Lambda(\chi_i^{\text{J}})[S_i^*]$ than the exact $J_i = \Lambda(\chi_i^{\text{J}})[S_i]$. Groups 1,2 and 4 show about 1 % deviation from the exact solution

above 0.5 Mm, and all mean intensities are accurate to within 5 % at all heights.

The first (uppermost) panel in Fig. 4 shows the deviations in radiative heating ϕ_i^* . The typical deviation is in the range 1-15 %, less than using J_λ^{PP} as a weight function (third panel). Since the mean intensities are lower, Eq. (27) in the streaming limit predicts a lower photon absorption rate, hence lower heating rate. Since the thermal emission rate $\kappa_i^{\text{B}} B_i$, and σ_i^{J} are exact in this case, lower J_i^* would decrease the heating.

This prediction is however not correct in all cases for the following reason. We have used Eq. (60) in the calculation of the flux divergence, and this would be equal to Eq. (27) only if the calculations were done analytically (in a continuous domain) and not on the discretized mesh. The $\Delta\tau_\Omega/\Delta r$ factor for the opacity in Eq. (60) is varying for different ray directions and not exactly equal to the opacity χ_i^{T} in Eq. (27). Hence, the second panel in Fig. 4 shows a lower heating than the exact solution only in group 1 and 2. Group 3 and 4 do not show the same behavior even if the mean intensities are lower also for these groups.

The middle panel in Fig. 5 show the effect on the ratio $\langle J_i^*/J_i \rangle$ for solutions using J_λ^{PP} as a weight function. These approximate solutions are on the average both higher and lower than the exact solutions. Hence, the plane parallel mean intensity as a weight function for the opacity coefficients introduces additional inaccuracies than just a reduction of the mean intensity in optically thin layers. As can be seen in the figure, the deviations have increased. All mean intensities are accurate to within ~ 15 % at all heights.

5.2. Convergence properties of the iteration method

At each timestep in a simulation, and for each group, we iterate the source function to convergence (as described in Sect. 4), and then the flux divergence is calculated according to Eq. (60). The initial estimate for the source function $S^{n=0}$ is taken as the converged source function from the previous timestep. We use Orthomin acceleration (Vinsome 1976) to improve the convergence rate of the iteration. We find that Orthomin acceleration (of order 2) reduces the number of iteration steps by a factor of about 2.

We stop the iteration when the maximum relative change

$$\text{Max} \left[\frac{|S^n - S^{n-1}|}{S^{n-1}} \right]$$

is smaller than a chosen limit, typically 10^{-3} (n is iteration number). This corresponds to typically 2, 5, 10 and 15 iterations for group 1, 2, 3 and 4 respectively. Increasing group number corresponds to smaller destruction probabilities in the atmosphere (see Fig. 1), and hence more iterations are needed.

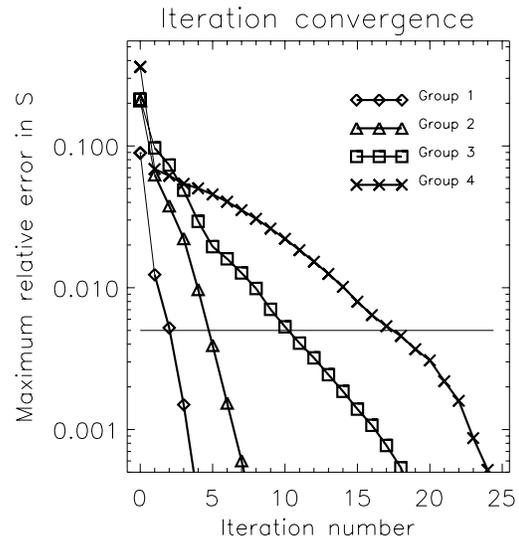


Fig. 6. Source function convergence for the iteration method, at an arbitrary timestep in a solar simulation. Maximum relative errors from the converged source functions are plotted as function of iteration number. The initial source functions are taken from the previous timestep. A relative error of 5×10^{-3} is marked by the line. This correspond to 2, 5, 10 and 17 iterations. These numbers are typical for the amount of iterations needed in simulations, in which we choose a convergence limit based on relative changes is the source function (typically 10^{-3}) between iterations. More iterations are needed for increasing group number, since the destruction probability (in the atmosphere) is smaller for increasing group number. The convergence curves in this log-plot are roughly linear, corresponding to an exponentially decreasing error with iteration number.

In Fig. 6, we have shown the maximum relative error or deviation from the converged source function S^∞ ,

$$\text{Max} \left[\frac{|S^n - S^\infty|}{S^\infty} \right],$$

as function of iteration number. We have chosen an arbitrary timestep in the simulation, and we have used the converged source function from the previous timestep as an initial estimate. In the figure, we see that the typical number of iterations given above (corresponding to 10^{-3} relative changes) correspond to errors of about 5×10^{-3} . Note that we choose a limit on the maximum relative change as a convergence criterion in simulations, since the fully converged source function is not calculated. From Fig. 6 we see that the errors decrease roughly exponentially with iteration number.

Errors at iteration number 0 show the maximum deviations between the converged source functions at current and previous timesteps. We see that these can be as large as 0.3–0.4 (group 4), even if the interval between timesteps are only on the order of 0.1 s of solar time. Iteration of the source function at every timestep is therefore needed.

6. Discussion

6.1. Comparison to previous LTE method

Here we compare the new method for scattering with the previous LTE method used in the convection simulations of Nordlund and Stein.

Nordlund's LTE scheme. We review this method here; additional information can be found in Nordlund (1982) and Nordlund & Dravins (1990). Nordlund treated the source function in LTE for all groups, $S_i^* = B_i$, and used a scaled opacity for the transport equation, $\chi_i^T = \chi_{i=1}^T 10^{i-1}$, where 10^{i-1} is the typical ratio between monochromatic opacities within group i and the opacity in group 1, at unity optical depth. The binning of wavelengths into groups satisfying these ratios at unity optical depth, was performed in the horizontally averaged model. Only $\chi_{i=1}^T$ and B_i were stored in the lookup table, and the flux divergence was calculated as in Eq. (27) with $S_i^* = B_i$.

Nordlund & Dravins (1990) calculated $\chi_{i=1}^T$ as a corrected standard opacity (Rosseland opacity integrated over continuum opacities only) to include the influence from the depth variation of weak iron lines in optically thin layers:

$$\chi_{i=1}^T = \chi_R^c(3D) X_{\text{corr}}.$$

$\chi_R^c(3D)$ is here the standard opacity, labeled to indicate a three dimensional structure. X_{corr} was calculated as a function of depth in the horizontally averaged model and weighted between diffusion and streaming regimes:

$$X_{\text{corr}} = \frac{\kappa_{i=1}^{\text{JPP}}}{\chi_R^c} e^{-2\tau_R^c} + \frac{\chi_R}{\chi_R^c} \left(1 - e^{-2\tau_R^c}\right),$$

$\kappa_{i=1}^{\text{JPP}}$ includes line plus continuum absorption opacities. The line opacities from the ODF were here treated as pure absorption opacities. χ_R^c is the standard opacity, τ_R^c is the standard optical depth, and χ_R the full Rosseland opacity, all quantities for the horizontally averaged model.

The flux divergence integral (Eq. 27) for group 1 takes this form in the streaming regime:

$$\chi_R^c(3D) \frac{\kappa_{i=1}^{\text{JPP}}}{\chi_R^c} (B_{i=1} - J_{i=1}^*),$$

where

$$\chi_R^c(3D) \frac{\kappa_{i=1}^{\text{JPP}}}{\chi_R^c}$$

is an approximation to the three dimensional absorption opacity. This approach approximately excludes the scattering contribution in the flux divergence integral. Note also that $J_{i=1}^*$ is still computed in LTE in this scheme, without iterations.

Details in absorption opacity was not accounted for in groups of higher numbers than 1. Scaling of the opacity in group 1 was assumed to be a sufficient approximation for these opacities. In the new method, we calculate all

group mean opacities explicitly, and we also include line scattering.

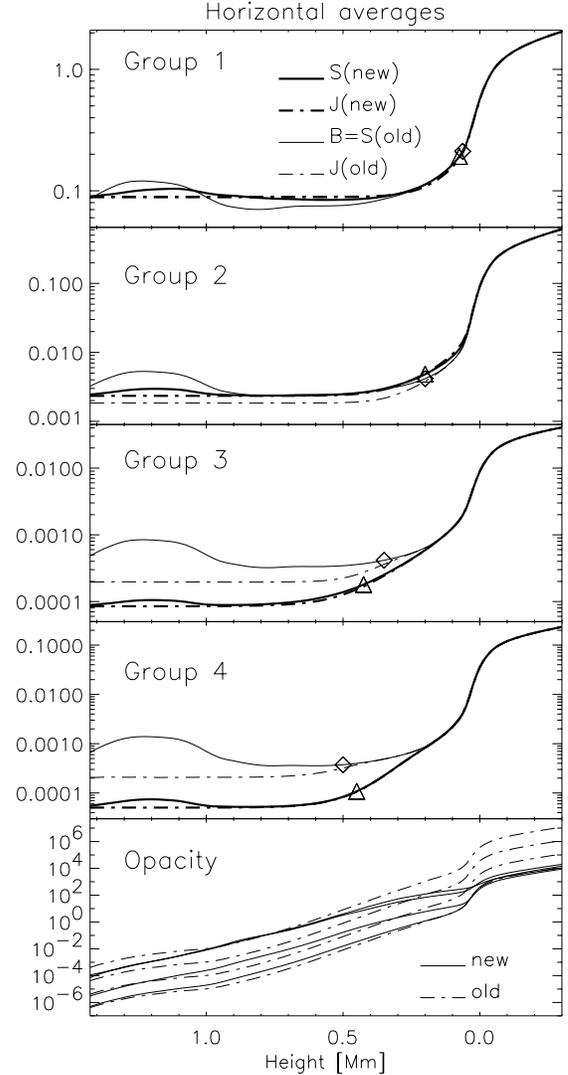


Fig. 7. Horizontally averaged group mean intensities J and source functions S from the test-snapshot. The units are arbitrary, but the scaling is the same for the four first figures. The mean intensities from the new method (new) are closer to the source functions than for the previous LTE method (old). The different values of mean intensities are mainly due to different source functions, but there is also some influence from differing heights where optical depths are unity. Diamonds and triangles show the heights for where the optical depth is unity. *Lower panel:* Horizontal averages of group mean opacities (arbitrary units). Higher values correspond to higher group number. Below 0.0 Mm, all groups have opacities equal to the group mean Rosseland opacity for the new method. Note the scaling by powers of ten for the “old” opacity.

Differences in mean intensity and flux divergence. Horizontal averages of mean intensities and source functions

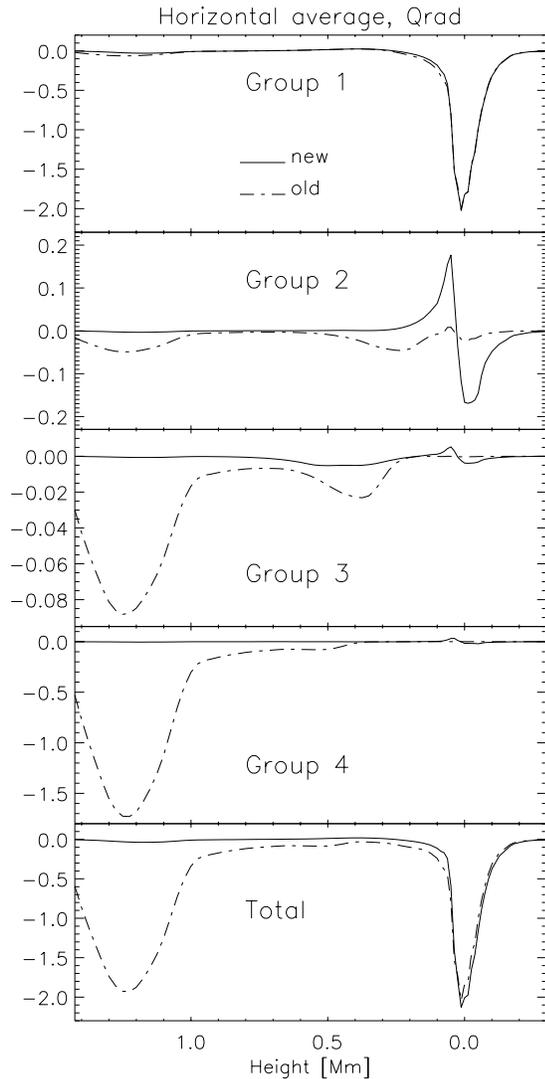


Fig. 8. Horizontal averages of radiative heating per mass unit. The units are arbitrary, and the scaling is the same for all figures. The amplitude of the flux divergence for the previous LTE method (old) is much larger above 0.3 Mm than for the current method. This effect is mainly due to the larger differences between the source function and the mean intensity. Around 0.0 Mm, the new method produces larger amplitudes in the sense that hot regions are cooled more and cooler regions are heated more. *Lower panel:* The sum of all groups. The main difference above 1.0 Mm comes from the contribution in Group 4.

are compared in Fig. 7 for the same snapshot that we have used previously (Fig. 1). We have used the same set of wavelengths per group for LTE and scattering results. The mean intensities from the new method are closer to the source functions than for the LTE method, due to the scattering contribution in the new method. Pure scattering would imply that $J_i^* = S_i^*$. The values of the mean intensities are also differing, mainly due to differences in the

source function, and partially due to the different heights for which the optical depths are unity.

Diamonds and triangles show the heights for where the horizontally averaged optical depth $\bar{\tau}_i$ is unity, for previous and current methods respectively. The mean intensities in optically thin regions are approximately $\langle J_i^* \rangle \simeq 1/2 \langle S_i^* \rangle (\bar{\tau}_i = 1)$. The group mean opacities that determines the optical depths are seen in the bottom panel of Fig. 7. The new opacities are higher than the old opacities above 0.3 Mm, except in Group 4, where the old opacity has been scaled to higher values over most of the height range. The reason for a lower opacity in group 1 for the LTE method is that scattering contributions to the opacity are excluded by the X_{corr} factor.

Radiative heating per mass unit is compared in Fig. 8. The heating amplitudes for the LTE method is much larger above 0.3 Mm than for the current method. The excessive cooling above 1.0 Mm is due to the temperature perturbation seen in Fig. 1. This effect is mainly due to larger difference between the source function and mean intensity. This larger difference is not compensated by the smaller LTE-opacities, such that the amplitudes in heating (cooling) become larger. We note that increased flux divergence amplitudes implies shorter radiative damping time of the waves that cause the temperature perturbations.

In the cooling layer around 0.0 Mm, the case is reversed. The new method produces larger amplitudes in the sense that hot regions are cooled more and cooler regions are heated more. This is because the mean intensities are smoother with scattering than without, hence the differences between local temperature and radiation temperature becomes larger. We should therefore expect smoother temperature gradients in the granular layer. Note that this effect occurs below optical depth unity for groups 2, 3 and 4, i.e., in relatively optically thick layers.

6.2. Solar simulations with and without scattering

We have also studied the effect of the two different methods on the chromospheric temperature structure. Two simulations were run, one with Nordlund's LTE-radiation method and one with the new scattering method. Both simulations were allowed to relax to a steady dynamic state in the chromosphere by running them for 0.5 hours of solar time. We used the same initial condition, and the same physical input parameters. The only difference was in the treatment of radiation. In the following, we compare the different terms in the gas-energy equation, where the radiative flux divergence enters, to explain the different temperature structures that results in upper atmospheric layers.

Average temperature structures. The mean temperature structures of the atmospheres are shown in the upper panel in Fig. 9 as full drawn lines. These are the aver-

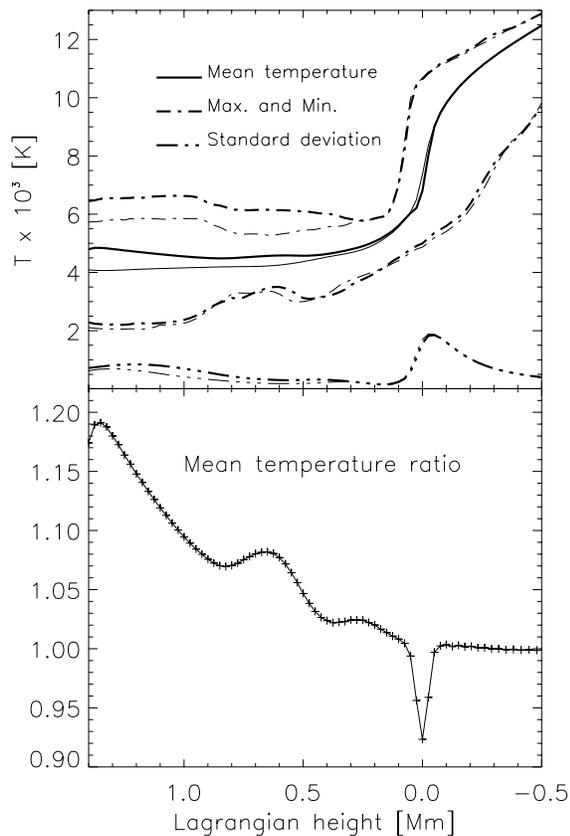


Fig. 9. *Upper panel:* Time averaged temperature, maximum and minimum values, and standard deviation. Thin curves: LTE, thick curves: scattering. The mean chromospheric temperature is higher (500 K) with scattering, and a marginal chromospheric temperature rise is seen. The temperature contrast (larger spread in temperature) and maximum temperatures are higher with scattering. *Lower panel:* Ratio (scattering/LTE) between the average temperatures. The spike at 0.0 Mm is due to the smoother temperature gradient with scattering. The Lagrangian scale follows the motion of the radial oscillations in the model, with typical periods of 3 min. in the chromosphere, and 5 min. in the convection zone.

age temperatures in time at the given height. The height scale is here “Lagrangian”, in the sense that we follow the average vertical oscillations, which have dominating amplitudes at 3 min. period in the chromosphere and 5 min. period in the convection zone. In this way, we eliminate advection of gas back and forth from these oscillations.

In chromospheric layers (0.5–1.4 Mm), the scattering solution (thick lines) shows a marginal temperature rise with height, while the LTE solution (thin line) does not. The scattering solution is typically 500 K hotter than the LTE solution. The minimum temperatures are approximately the same (close to 2000 K in upper layers), but the maximum temperatures are typically 1000 K higher for the scattering solution, and reaching 6500 K. Both high temperatures (6500 K) and low temperatures (2000 K) are

coexisting at a given time and height in the simulation. As seen from the standard deviation, the scattering solution produces larger temperature variations.

In the cooling layer at 0.0 Mm, the scattering solution gives a smoother temperature gradient. This is also shown in the lower panel in Fig. 9 which shows the ratio between average temperature from the scattering solution and average temperature from the LTE solution.

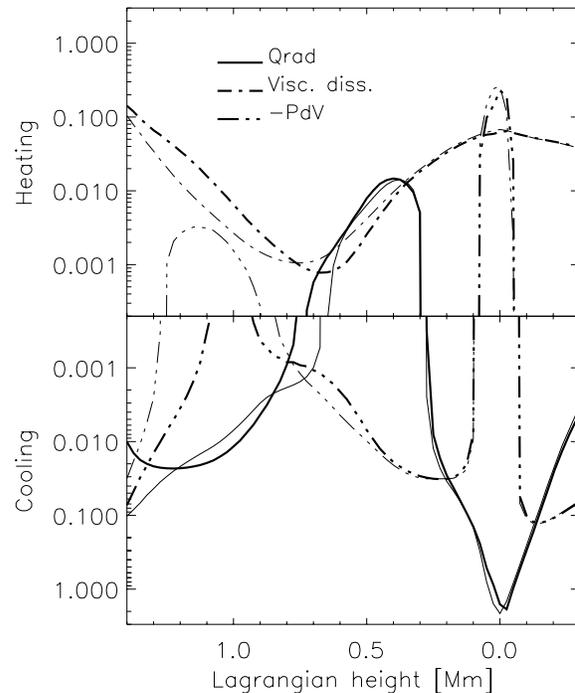


Fig. 10. Time averaged terms in the gas-energy equation. All terms have been scaled with the same factor, and the units are dimensionless. Thin curves: LTE, thick curves: scattering. Radiative cooling almost balances viscous dissipation in the range 0.8–1.2 Mm. The P-dV work is relatively smaller than these terms in this range. Below 0.5 Mm, the solutions are quite similar. The Lagrangian scale follows the motion of the average oscillations, with typical periods of 3 min. in the chromosphere, and 5 min. in the convection zone.

Energy balance in the chromosphere. What causes the differences in temperature distribution and mean temperature in the chromosphere? First, we consider the gas-energy equation (first law of thermodynamics):

$$\frac{De}{Dt} = \Phi + Q_{\text{rad}} - \frac{P}{\rho} \nabla \cdot \mathbf{u},$$

where e is the specific internal energy, Φ is dissipation due to viscosity, $Q_{\text{rad}} = -\nabla \cdot \mathbf{F}_{\text{R}}/\rho$ is specific radiative heating, and $-(P/\rho)\nabla \cdot \mathbf{u}$ is mechanical work done on the gas (“P-dV” work per gram). We take the horizontal average and time average of this equation on the oscillating

Lagrangian scale. The time averaging is done over an integer number of periods of the dominating amplitude. An atmosphere in steady state implies that the internal energy (on the oscillating Lagrangian scale) averaged over a limited time interval stays constant. In atmospheric layers where the motions are oscillatory, this results in a vanishing $\langle De/Dt \rangle$, where $\langle \rangle$ denotes averaging. In photospheric layers, we get a non-zero contribution from $\langle \mathbf{u} \cdot \nabla e \rangle$ by convective overshoot flow.

In chromospheric layers (0.8–1.2 Mm), the net P-dV work over a cycle tends to be relatively small, with one half cycle of the oscillation cancelling the other half cycle. The time average is seen in Fig. 10. Hence, we get essentially a balancing between viscous dissipation (always positive) and radiative cooling,

$$\langle \Phi \rangle \simeq -\langle Q_{\text{rad}} \rangle.$$

The mean temperature and temperature distribution at a given height in the *optically thin* chromosphere is controlled by the density and temperature dependency of radiative heating, hereafter referred to as the heating function $q_{\text{rad}}(\rho, T)$. The definition of this heating function is only meaningful in optically thin layers where the mean intensity has small relative spatial and temporal fluctuations. We can therefore treat the mean intensity as a constant with a well defined radiation temperature, such that the absorption of photon energy is locally determined by the density and temperature dependency of the absorption opacity. Emission of photon energy is in our case locally determined everywhere in the atmosphere through the absorption opacity and Planck function.

We stress that the radiation field in optically thin layers is determined from the radiation transport with scattering in optically thicker layers below. In these layers, it is not possible to define a heating function, since the fluctuations in mean intensity is not relatively small. We will therefore only discuss the energy balance of the chromosphere in layers where the heating function can be defined.

The energy source from viscous dissipation is balanced by radiative cooling, such that over a limited time interval, the temperature probability distribution $p(z, T)$ at a given height satisfies this (approximate) relation:

$$\langle \Phi(z) \rangle \simeq - \int_T q_{\text{rad}}(\langle \rho(z) \rangle, T) p(z, T) dT. \quad (61)$$

Here, we have used the horizontal average of the density at the given height to represent the average heating function at a given temperature. This also provides a single valued function with temperature as variable. The heating function has different functionality for different groups. The general behavior is that the absorption coefficient increases with increasing temperature, as does the Planck function. This results in a steeper slope in the negative part (the cooling region) where the local gas temperature is higher than the radiation temperature.

Fig. 11 shows the chromospheric heating functions in LTE (Nordlund's approach) in the upper panels, and

Fig. 12 shows the heating functions with scattering. These functions are averages over $q_{\text{rad}}(\langle \rho(z) \rangle, T)$ over the height interval 0.8–1.2 Mm to show typical behavior. This is meaningful, since the heating functions are normalized by density, and is therefore comparable in magnitude. Furthermore, they have similar functionality in this height range.

The bottom panels in these figures show the temperature probability distributions as full drawn lines with crosses, and they are sampled from the same height interval and over the full duration of the simulation. Vertical lines mark the average temperature in this height interval. The horizontal lines in the middle panels show the average radiative cooling (equal to the magnitude of viscous dissipation in this range as seen in Fig. 10) that results when these temperature distributions are integrated over the heating functions according to Eq. (61).

We see that with scattering, the amplitude and slope in the cooling region (for the same temperature) are smaller than in LTE (the logarithmic slope in the plotted heating functions reflects the *relative* linear derivative). The compressional parts of the waves cool by radiation (parts of the waves with positive temperature perturbations corresponding to the cooling region). With scattering, these parts of the waves must reach higher temperatures before radiative cooling can balance viscous dissipation. The temperature distribution function is therefore broader with scattering, due to the tail towards higher temperatures up to 6500 K. The distribution for temperatures lower than 4000 K is similar to the LTE solution. Hence, the mean temperature is higher with scattering due to this broader distribution in the high temperature range.

As can be seen in these figures, the cooling is dominated by different groups for the two cases. The group for the strongest spectral lines dominates in LTE (group 4 in Fig. 11), while with scattering, the group for continua and the weakest spectral lines (weak iron lines) dominate (group 1 in Fig. 12).

The distribution marked with a dash-dotted line in Fig. 11, is the temperature probability distribution for the initial atmosphere for the LTE-run. This was taken from previous scattering runs. The temperature probability distribution (full drawn) for the complete run shows that the atmosphere has cooled down and settled at a narrower distribution, satisfying the rate of viscous dissipation.

In Fig. 10 we also see that the scattering solution dissipates more (thick line) than the LTE solution (thin line) in the chromosphere. This can be due to less wave damping by radiation, allowing more wave energy flux to reach the chromosphere from the convection zone. We note that the same viscosity was applied to both simulations, so the influence from viscous work (always negative) on the damping of wave energy flux was given the same treatment.

Despite of different wave damping, we can conclude that given the same amount of viscous dissipation in the

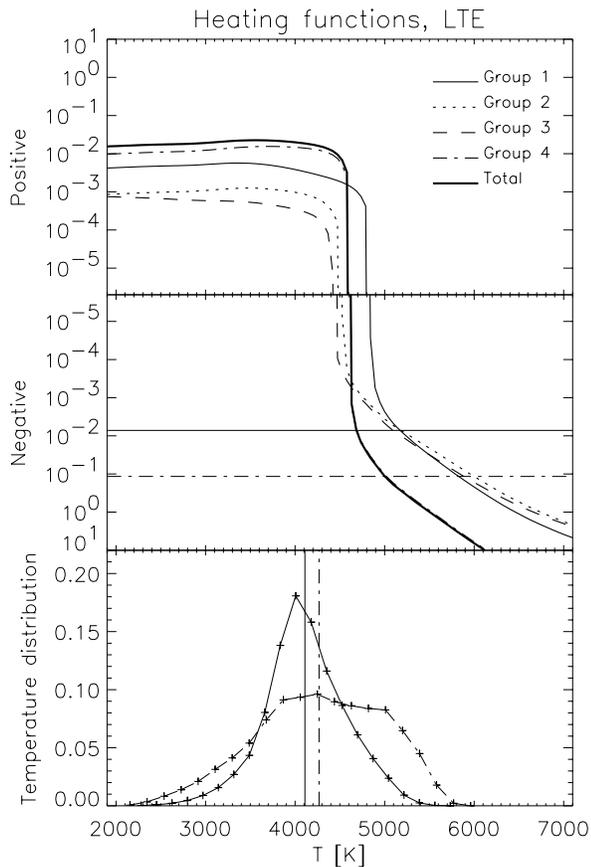


Fig. 11. *Upper panels:* Chromospheric heating functions in LTE (the units are dimensionless, and the scaling is the same as in the gas-energy equation). These are averaged in the height interval between 0.8 Mm and 1.2 Mm to show typical behavior. The full drawn horizontal line shows the time averaged (steady state) radiative cooling in the same region. The dash-dotted line shows the radiative cooling in the initial atmosphere for this LTE simulation. The snapshot was taken from a scattering run. The initial cooling is larger than the steady state value, showing an initial loss of energy from the chromosphere. *Lower panel:* Temperature probability distributions in the same region. Full drawn line: steady state distribution. Dash dotted line: initial distribution. Vertical lines: average temperatures. The distribution is narrower and the mean has decreased in steady state. The integral over the heating function with the temperature probability distribution as a weight function gives the time averaged heating (which is negative in both cases).

optically thin chromosphere, the maximum temperatures and temperature contrast (width of the distribution) will be larger with scattering than in LTE.

Energy balance in the photosphere. In photospheric layers from 0.0 Mm to 0.5 Mm, the two solutions do not differ very much as seen in Fig. 10. P-dV work is done on the gas at 0.0 Mm as it compresses in the cooling layer when it loses internal energy to radiation. Here we encounter

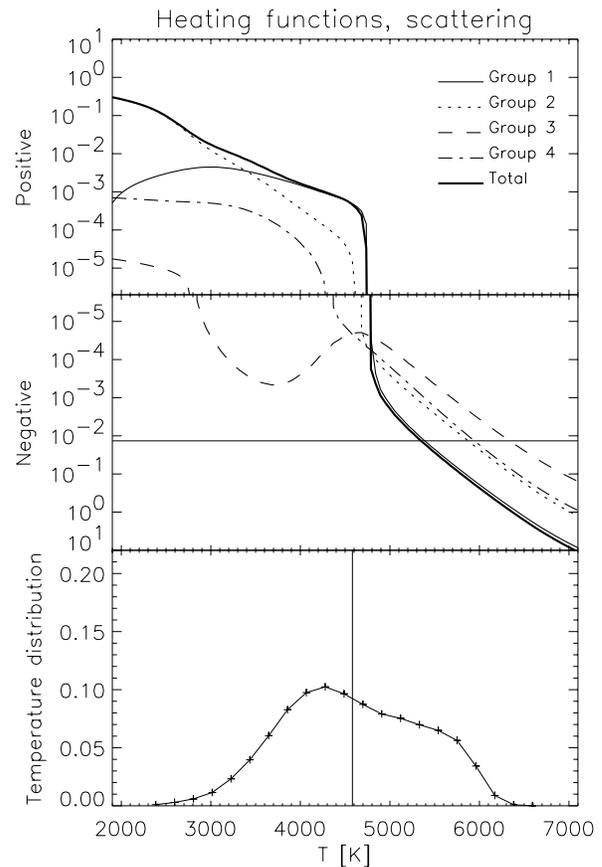


Fig. 12. *Upper panels:* Chromospheric heating functions with scattering included (the units are dimensionless, and the scaling is the same as in the gas-energy equation). These are also averaged in the height interval between 0.8 Mm and 1.2 Mm. The full drawn horizontal line shows the time averaged (steady state) radiative cooling that balances viscous dissipation. *Lower panel:* Temperature probability distribution in steady state. Vertical line: average temperature. The distribution is wider than in LTE with larger maximum temperatures. Temperatures corresponding to cooling serves to balance the energy input from viscous dissipation. Since the magnitude of the heating function is less than in LTE, temperatures are higher to balance viscous dissipation. Even with the same amount of dissipation as in LTE, the temperatures would be higher with the scattering-cooling function to balance the dissipation rate.

a large influence from advection and the radiative cooling is larger than the two other terms (since we do not follow the gas parcels, only the oscillatory motion). Between 0.3 Mm and 0.5 Mm, the gas expands over granules, and contracts over downdrafts, but the net horizontal average is expansion work that is compensated by radiative heating.

The smoothing of the temperature gradient in the cooling layer with scattering as seen in Fig. 9, is explained by radiative transfer in groups 2, 3 and 4. In optically

thick layers ($\tau > 1$), the mean intensities are in general smoother than the LTE solution. This serves to increase the difference between the Planck functions and the mean intensities, and therefore energy is transported more efficiently from hot to cool regions, as long as the mean photon destruction probability is not too small to compensate for this effect. As a result, temperature gradients in the cooling layer tend to smooth out.

The mean intensities in these groups drop below the source function in deeper convective layers than in the LTE case, and therefore the cooling region extends to deeper layers (Fig. 8). Immediately above 0.0 Mm, (~ 50 km) the mean intensities are higher than the source function, and heating occurs in the same groups. The total radiative cooling profile of the horizontal average is therefore shifted slightly downwards.

7. Summary

The goal of this work is to develop numerical methods for an approximate solution of the radiative flux divergence in 3D simulations of stellar atmospheres. This method is designed to handle upper atmospheric layers where photon mean free paths are large and photon scattering can be important, and deeper atmospheric layers where photon transport is diffusive. To make this problem tractable, we have made three fundamental approximations: 1) opacities are calculated in LTE, 2) group mean opacities (e.g., Mihalas & Mihalas 1984) substitutes the monochromatic spectrum, and 3) coherent and isotropic photon scattering in the source function.

Using these approximations, we calculate group mean scattering, absorption and total opacities by wavelength integration of the transport equation. The opacities have different definitions for optically thin and thick layers. In optically thin layers, we use an approximate radiation field (in the solar case, the one dimensional radiation field) as weighting function, except for the group mean absorption opacity for which we use the exact Planck average. In optically thick layers we use Planck averages, except for the total opacity. Here we use Rosseland opacities such that the diffusion solution of the transport equation is satisfied. The group mean source function is, by the usual definition, the ratio between emissivity and total opacity. This source function contains an approximate scattering term and an exact contribution from thermal emissivity.

This approach results in a three dimensional scattering problem in each group equivalent to a Fredholm integral equation of the second kind for the source function, or an integro-differential equation in bi-directional intensity. This problem is solved by iteration using a new method based on a one-ray approximation in the angle integral for the mean intensity. The equations to be iterated are tri-diagonal matrix equations, one for each ray. For this coherent problem, the method is faster in terms of computer time than the local ALI method (Cannon 1973b,

Olson et al. 1986 and Hubeny 1992), Nordlund's method (Nordlund 1985), and Cannon's AQPT method (Cannon 1973a).

We have tested the methods in solar convection-atmosphere simulations (Skartlien et al. 1998), and compared exact wavelength integrated monochromatic solutions with the corresponding approximate group mean solutions. We find that the total flux divergence obtained from groups deviates with less than 10 % from the exact solution, with maximum deviation in atmospheric layers. Individual groups that transport less energy deviate with typically 30 % in atmospheric layers.

Solutions of the source functions are needed every timestep in simulations. The initial source functions for the iteration procedure are taken as the converged source functions from the previous timestep, and these can deviate with more than 10 % from the converged source function at the current timestep. Iterations are therefore needed at every timestep to obtain a better accuracy. The time increment between two adjacent timesteps is typically 0.1 s of solar time. In these simulations, we need typically 2, 5, 10 and 15 iterations for the four groups we have used, to reach an accuracy of about 5×10^{-3} in the source functions. The CPU-time used for solving for the radiation scales linearly with the number of angles, gridpoints and wavelength groups.

We have compared Nordlund's LTE method (Nordlund 1982 and Nordlund & Dravins 1990) that has been used in solar and stellar convection simulations, with the current scattering method. The LTE method gives larger amplitudes of radiative heating/cooling in upper atmospheric layers, given the same temperature perturbations. This is mainly due to larger differences between the source function (Planck function in LTE) and mean intensity.

Tests with solar simulations show how the steady state structure of the atmosphere depends on the treatment of radiation. In lower atmospheric layers, and in the cooling layer in the convection zone, the differences are marginal. In upper atmospheric layers, the differences can not be neglected. In LTE (Nordlund's approach), the temperature probability distribution is narrower at a given height, and the mean temperature is lower than with scattering. The maximum chromospheric temperatures are about 1000 K higher with scattering, reaching 6500 K. The minimum temperatures in the chromosphere are about the same (2000 K).

The differences in temperature probability distributions in *optically thin* chromospheric layers are explained in terms of the different amplitudes of radiative heating at the same temperature. We find that energy input from (the same amount of) viscous dissipation is balanced by radiative cooling at higher temperatures with scattering. The simulation with scattering also allows for larger temperature perturbations and wave amplitudes. The dissipation is therefore slightly higher with scattering.

Clearly, improvements can be done for the methods described in this paper, such as inclusion of UV-opacity from iron lines in the ODF for spectral lines, more accurate treatment of destruction probability than the Van Regemorter (1962) formula, and a more accurate description than isotropic scattering in the source function. A detailed comparison to 1D simulations treating fully consistent NLTE effects in chromospheric layers would be useful since NLTE spectral line radiation is important in the energy balance in these layers (e.g., Carlsson & Stein 1992). An inclusion of the effect from radiation on opacities could possibly be done in the near future using approximate NLTE line blanketing methods developed in Anderson (1985, 1989).

Acknowledgements. This work was supported by a PhD-grant, and through a grant of computing time (Program for super-computing), from the Research Council of Norway. Andreas Botnen is thanked for comparisons with his 3D Short Characteristics method and for help with FORTRAN 90 and MPI-programming. Mats Carlsson and Teresa Palmer are thanked for reading the manuscript carefully. Colorado Research Associates (CoRA) is thanked for hospitality and for providing additional computer resources during this work.

References

- Anderson L. S. 1985, *ApJ*, 298, 848
 Anderson L. S. 1989, *ApJ*, 339, 558
 Auer L. H., Bendicho P. F., Bueno J. T. 1994, *A&A*, 292, 599
 Botnen A. 1997, Radiative transfer in 3D, Cand. Scient. Thesis, Inst. Theor. Astrophys. Oslo
 Cannon C. J. 1973a, *J. Quant. Spectrosc. Radiat. Transfer*, 13, 627
 Cannon C. J. 1973b, *ApJ*, 185, 621
 Cannon C. J. 1976, *A&A*, 52, 337
 Carlsson M. 1986, A Computer Program for Solving Multi-Level Non-LTE Radiative Transfer Problems in Moving or Static Atmospheres, Report No. 33, Uppsala Astronomical Observatory
 Carlsson M., Stein R. F. 1992, *ApJ*, 397, L59
 Carlsson M., Stein R. F. 1997, *ApJ*, 481, 500
 Castor J. I., Dykema P. G., Klein R. I. 1991, in L. Crivellari, I. Hubeny, D. G. Hummer (eds.), *Stellar Atmospheres: Beyond Classical Models*, NATO ASI Series C 341, Kluwer Academic Publishers, Dordrecht, p. 49
 Feautrier P. 1964, *Comptes Rendus Acad. Sci. Paris*, 258, 3189
 Gustafsson B. 1973, *Uppsala Astr. Obs. Ann.*, 5, No. 6
 Gustafsson B., Bell R. A., Eriksson K., Nordlund Å. 1975, *A&A*, 42, 407
 Hauschildt P., Störzer H., Baron E. 1994, *J. Quant. Spectrosc. Radiat. Transfer*, 51, 875
 Holweger H. 1970, *A&A*, 4, 11
 Hubeny I. 1992, in U. Heber, C. S. Jeffery (eds.), *The Atmospheres of Early-Type Stars*, Lecture Notes in Physics 401, Springer, Berlin, p. 377
 Jacobi C. G. 1845, *Astronom. Nachr.*, 32, 297
 Kislman D., Nordlund A. 1995, *A&A*, 302, 578
 Klein R. I., Castor J. I., Greenbaum A., Taylor D., Dykema P. 1989, *J. Quant. Spectrosc. Radiat. Transfer*, 41, 199
 Kunasz P. B., Auer L. H. 1988, *J. Quant. Spectrosc. Radiat. Transfer*, 39, 67
 Kunasz P. B., Olson G. L. 1988, *J. Quant. Spectrosc. Radiat. Transfer*, 39, 1
 Magain P. 1983, *A&A*, 122, 225
 Mason H. E., Monsignori-Fossi B. C. 1994, *A&AR*, 6, 123
 Mihalas D. 1978, *Stellar Atmospheres*, W. H. Freeman and Co., San Francisco (second edition)
 Mihalas D., Auer L. H., Mihalas B. R. 1978, *ApJ*, 220, 1001
 Mihalas D., Mihalas B. W. 1984, *Foundations of Radiation Hydrodynamics*, Oxford Univ. Press, New York
 Nordlund Å. 1982, *A&A*, 107, 1
 Nordlund Å. 1985, in J. E. Beckman, L. Crivellari (eds.), *Progress in stellar spectral line formation theory*, NATO ASI Series C152, Reidel, Dordrecht, p. 215
 Nordlund Å. 1991, in L. Crivellari, I. Hubeny, D. G. Hummer (eds.), *Stellar Atmospheres: Beyond Classical Models*, NATO ASI Series C 341, Kluwer, Dordrecht, p. 61
 Nordlund Å., Dravins D. 1990, *A&A*, 228, 155
 Nordlund Å., Stein R. F. 1990, *Comp. Phys. Comm.*, 59, 119
 Nordlund Å., Stein R. F. 1991, in L. Crivellari, I. Hubeny, D. G. Hummer (eds.), *Stellar Atmospheres: Beyond Classical Models*, NATO ASI Series C 341, Kluwer, Dordrecht, p. 263
 Olson G. L., Auer L. H., Buchler J. R. 1986, *J. Quant. Spectrosc. Radiat. Transfer*, 35, 431
 Olson G. L., Kunasz P. B. 1987, *J. Quant. Spectrosc. Radiat. Transfer*, 38, 325
 Pomraning G. C. 1971, *J. Quant. Spectrosc. Radiat. Transfer*, 11, 597
 Puls J., Herrero A. 1988, *A&A*, 204, 219
 Rosseland S. 1924, *MNRAS*, 1924, 525
 Rybicki G. B. 1984, in W. Kalkofen (ed.), *Methods in Radiative Transfer*, Cambridge Univ. Press, Cambridge UK
 Scharmer G. B. 1981, *ApJ*, 249, 720
 Skartlien R., Stein R. F., Nordlund Å. 1998, *A&A*, in preparation
 Spruit H. C., Nordlund Å., Title A. M. 1990, *ARA&A*, 28, 263
 Steffen M. 1990, *A&A*, 239, 443
 Stein R. F., Nordlund Å. 1989, *ApJ*, 342, L95
 Uus U. 1987, *Tartu Astrofüüs. Obs. Publ.*, 52, 33
 Van Regemorter H. 1962, *ApJ*, 136, 906
 Våth H. M. 1994, *A&A*, 84, 319
 Vinsome P. K. W. 1976, in *Proceedings of the 4th Symposium on Reservoir Simulation*, Soc. of Petroleum Engineers, Boulder, p. 149

Paper II

EXCITATION OF WAVE TRANSIENTS BY SOLAR GRANULATION

R. Skartlien ¹

¹ Institute of Theoretical Astrophysics, P.O. Box 1029, Blindern, N-0315 Oslo, Norway

Abstract

Results from 3D numerical simulations of the non-magnetic solar atmosphere and the upper convection zone are presented. Transient acoustic wave trains in the atmosphere are excited at the top of the convective zone (the cooling layer) and immediately above in the convective overshoot zone, by small granules that undergo a rapid collapse, in the sense that upflow reverses to downflow, on a timescale shorter than the atmospheric acoustic cutoff period (3 min). An extended downdraft between larger cells is formed at the site of the collapse.

The waves produced are long wavelength, gravity modified acoustic waves with periods close to the 3 minute eigen-period (or cutoff period) of the solar atmosphere. The oscillation is initially horizontally localized with a size of about 1 Mm. The wave amplitude decays in time as energy is transported horizontally and vertically away from the site of the event. Granule collapse is the only convective wave source I find that is responsible for a clear cause and effect relationship between the convection zone and atmosphere.

Observed darkening of intergranular lanes and the associated photospheric wave motions, the so called "acoustic events" (Restaino et al. 1993, Rimmele et al. 1995 and Espagnet et al. 1996), could be explained by this purely hydrodynamical process. Furthermore, the observed "internetwork bright grains" in the CaII H and K line cores and associated shock wave trains in the chromosphere (Rutten & Uitenbroek 1991, Lites et al. 1993, Carlsson & Stein 1994, 1997, and Cheng & Yi 1996), can also be linked to the wave transient.

Keywords: Convection – Hydrodynamics – Wave excitation – Sun: atmosphere – Sun: granulation – Sun: oscillations

1 Introduction

This work aims at a better understanding of the generation of *non-magnetic* acoustic waves in the solar convection zone, and how these waves propagate into, and interact with, the overlying atmosphere. I am mainly concerned with generation of waves that have a large impact on chromospheric dynamics.

The solar convection zone is turbulent in its interior. Non-adiabatic cooling and heating by radiation occurs in the upper layer of the convection zone. It is well known from the theory of acoustics that turbulence and non-adiabatic effects can produce sound waves in a compressible medium. The question is therefore not whether waves can be produced in the convection zone, but instead *how* they are produced and how they interact with the atmosphere above.

Physics regarding wave sources and waves can not be observed directly because the radiation we receive contains information averaged over large volumes, and physical quantities like temperature and velocity do not translate directly to observables. In addition, the interior of the convection zone, where wave sources can operate, can not be observed at all. Numerical simulations are therefore needed to resolve these questions.

Simulations of the solar atmosphere have been done by several authors (e.g., Rammacher & Ulmschneider 1992, Carlsson & Stein 1992, Cheng & Yi 1996 and Carlsson & Stein 1997) in plane parallel geometry. These models can not treat convection dynamics, and wave generators in the convection zone can therefore not be taken into account in the same models.

Three dimensional convection-atmosphere models are needed to understand which convective processes can significantly influence atmospheric dynamics. The upper convective layers and the convectively stable atmosphere should be modelled simultaneously. In the current work, I use fully compressible 3D simulations in which soundwaves and their convective sources operate simultaneously and consistently. Also the possible back reaction from waves to sources (and convective flow) is therefore automatically accounted for. The convection model of Nordlund and Stein (hereafter NS) in its *non-magnetic* form is used as

a basis for this study. For a description of this model, see Nordlund & Stein (1990).

I have extended this model upwards into chromospheric layers. Photon scattering is equally important as photon absorption in the chromosphere, and that has called for an improvement in the treatment of radiation. The radiation method I have developed for the current work treats photon scattering approximately, and is described in Skartlien (1998).

1.1 Turbulent wave excitation in the convection zone

In classical acoustics, excitation of sound waves has been based on the theory of Lighthill (1952). Turbulent Reynolds stresses (e.g., in a jet-stream) and non-adiabatic processes (e.g., a heat source) produce waves. Generation of acoustic waves in the solar convection zone has been studied by e.g., Stein (1967), Goldreich & Kumar (1990) and Musielak et al. (1994). Most of these theories have assumed adiabatic conditions. The approach has been to combine a prescribed turbulence spectrum (e.g. a Kolmogorov spectrum) with Lighthill's theory of sound generation to make a statistical prediction of the wave flux.

In the Lighthill procedure, the governing equations are manipulated to form a linear wave operator on one side of the equality sign, while the remaining nonlinear terms on the other side are interpreted as source terms for the waves. With non-vanishing sources, this wave equation is driven, and for vanishing sources, it governs free wave propagation.

The standard procedure is to solve for the sound wave field at large distances from the source. The source itself is assumed to be localized volumes of turbulence. The solar convection zone is highly turbulent with most of the eddies confined to the narrow, compressed downflows. The broader, expanding upflows are essentially turbulence free in comparison to the downflows. Thus, the assumption of a localized source can be justified.

The Reynolds stresses arising in the source term have been treated by many authors in terms of vorticity as in Howe (1975) and Crighton (1981). The convection model used in this work shows that vortic-

ity is baroclinically generated¹ at the edges of granular cells, where isosurfaces of density cross isosurfaces of pressure (non-vanishing $\nabla\rho \times \nabla P$). At the edges of the cells I find vortex rings. Between down-drafts and cells I find vortex tubes and sheets, formed by the shear between downflow and upflow. These vortex structures are intensified (in terms of vorticity magnitude $|\vec{\omega}| = |\nabla \times \vec{u}|$) by the horizontal compression and vertical stretching in the downdrafts. The vorticity magnitude increases down to about 100 km below the cooling layer (measured in NS convection simulations with $253 \times 253 \times 168$ gridpoints covering $6 \times 6 \times 2.5$ Mm). Below that level, dissipation serves to decrease the magnitude. Thus, the most efficient zone of acoustic generation from turbulence is located near 100 km below the cooling layer.

Some of the turbulence generated acoustic energy will enter the atmosphere above the convection zone. The random location of the sources produces an interfering soundwave pattern that comprises randomness and no preferred spatial positions, at least in the higher layers of the atmosphere. We would also expect to observe this with a stationary signature.

1.2 Granular wave excitation

The low frequency ‘‘acoustic events’’ observed by Rimmele et al. (1995) and Espagnet et al. (1996) show, in contrast, spatially localized wave transients that stand out clearly from the random background field, and I suggest that their sources should be sought in processes other than turbulence within the convection zone.

The observed localized brightenings in the cores of the CaII H and K lines, with horizontal extent of about 1-2 Mm (the ‘‘bright grains’’), strengthens this view. The bright grains (review by Rutten & Uitenbroek 1991) imply vertically propagating chromospheric shockwaves that originate from lower amplitude waves in the photosphere (Carlsson & Stein 1997). They show a transient signature with 2-5 consecutive grains occurring at the same horizontal position.

Such large scale, spatially coherent waves might be excited by the continuously evolving granular pat-

¹ The vorticity equation is (with $\vec{\omega} = \nabla \times \vec{u}$ and \vec{u} is velocity):

$$\frac{\partial \vec{\omega}}{\partial t} = \nabla \times (\vec{u} \times \vec{\omega}) + (1/\rho^2) \nabla \rho \times \nabla P + \text{Viscous diffusion}$$

The first term is the advection term. The strength of the vorticity field increases by convergent flow perpendicular to the fieldlines and by divergent flow along the fieldlines. The second term is the baroclinic source, that generates vorticity (e.g., the torque produced by a density perturbation that is not in balance with the pressure gradient in a stratified medium). The third term represents a sink due to viscous diffusion of fieldlines. Note the analogy to the magnetic induction equation (excluding the baroclinic source).

tern. The efficiency of wave generation by the granules is determined by the timescale they evolve on. Rapid changes, with a timescale smaller than the atmospheric cutoff period, are efficient acoustic sources, while slow events, with timescales larger than the buoyancy period, are potential gravity wave sources.

The granular pattern at the top of the convection zone is a result of mass conservation and the balance between upward transportation of internal energy and radiative cooling. An imbalance here results in a change of granular topology, and an accompanying adjustment of pressure to restore a stable configuration. Excitation of pressure waves is thus expected.

The topological changes have a typical pattern. As a cell expands, the upflow velocity in the central parts is reduced (see Rast 1995). The reduced upward internal energy flux comes out of balance with the surface radiative cooling, and the temperature decreases. The lower temperature of the cooled gas has reduced opacity (mainly due to H^- opacity that is proportional to $\sim T^8$) and becomes transparent, such that the underlying gas also cools radiatively. The cooling stops when the gas temperature reaches the radiation temperature from the surrounding medium. As the gas cools, its density increases, and brakes the upflow further.

This “cooling erosion” of a granule splits it into several smaller fragments (e.g., Nordlund 1985). The largest of these fragments have high enough upward energy flux and expand until they finally undergo a new splitting process. The smaller fragments diminish in size until they finally disappear into neighboring downdrafts. I find that these smaller fragments disappear on timescales that are small enough to excite atmospheric waves in the acoustic domain. The granule fragmentation itself and “exploding granules”, which is the collapse in the central parts of a large granule, are processes that are not rapid enough to produce propagating acoustic waves. Observations indicate that wave excitation events, the so-called “acoustic events”, are found near intergranular lanes (Rimmele et al. 1995). These events are therefore consistent with the picture of small granule fragments that vanish in downdrafts while generating waves.

In addition to granule splitting, the surface cooling produces dense volumes of gas that create downward moving plumes² that penetrate into the convection zone. Supersonic plume velocities can result within the convection zone, until the gas finally ionizes and becomes buoyant (Rast et al. 1993 and Rast 1997). Accompanying large pressure fluctuations generate acoustic waves, and can possibly also excite p-modes

²The NS simulations show that the vortex structure in a plume is characterized by an expanding, horizontal vortex ring at the plume head. First, the plume head descends rapidly, then it slows down while the vortex ring expands.

(Rast & Bogdan 1998). This work shows that plumes in the convection zone that are produced by vanishing granules, are not the primary wave sources responsible for generation of *atmospheric* acoustic waves. Atmospheric waves generated by vanishing granules are instead excited in the convective overshoot zone, including the cooling layer.

The organization of the paper is as follows. I give an overview of the simulation model and describe some diagnostic tools in Sect. 2. The simulation results, regarding the collapse of small granules and the corresponding wave excitation, are discussed in Sect. 3. I draw the conclusions in Sect. 4.

2 Methods

I give a brief description of the numerical model in Sect. 2.1. A few diagnostic tools that are used to analyze the model results are described in Sect. 2.2, together with background physics needed for the analysis of the wave excitation events.

2.1 Numerical simulation

I have extended the solar convection model of Nordlund & Stein (1990) vertically to include chromospheric layers. The previous version of the code included the convective overshoot zone in the photosphere, and the upper boundary was located at about 500 km above the surface (cooling layer) of the convection zone. The new model extends to 1500 km above the top of the convection zone. This covers the photosphere and a large part of the chromosphere. The lower boundary is located 1500 km below the top of the convection zone. This gives a total vertical extent of 3 Mm, covering two orders of magnitude in density in the convection zone, and five orders of magnitude in the atmosphere. The density at the transition between convective layers and the convectively stable atmosphere is about 10^{-7} g cm⁻³. The horizontal extent is 6×6 Mm, which covers typically five large granular cells.

This model enables me to study the dynamic coupling between the convection zone and the atmosphere above. So far, I have excluded magnetic fields, which is a good approximation for the internetwork regions with relatively weak magnetic fields. This approximation is no longer valid in layers above 1500 km, as we approach the transition zone to the solar corona where the Lorentz force is of equal importance as gravity and pressure gradient forces. The ratio of gas pressure/magnetic pressure varies from $\sim 10^4$ in the internetwork photosphere to $\sim 10^2$ in the internetwork chromosphere and to $\sim 10^{-4}$ in the corona.

The inclusion of the dilute upper atmosphere required a new treatment of photon transfer to account for radiative heating/cooling in the energy balance. It is only in the energy balance that radiation enters for solar conditions. The new radiation method is described in (Skartlien 1998).

Briefly, the new method includes scattering processes between photons and ions. The importance of photon scattering is twofold. First, large gas volumes of the atmosphere are coupled in the sense that thermal information can propagate (with the speed of light) via photons in every direction. Photons may be re-scattered or absorbed as they arrive at a certain location. The local gas temperature is coupled to the mean radiation temperature by absorption of non-local photons. Second, if scattering processes dominate over absorption processes, then the coupling of local gas to the radiation field is weaker. This is often the case in the upper atmospheric layers where photons are mainly scattered until they escape into space. The coupling is stronger in deeper atmospheric layers.

The interaction between radiation and gas-dynamics works both ways. Radiation modifies the dynamics through heating or cooling, and dynamics modifies the radiation field by changes in the spatial opacity distribution. This coupling is important both in the upper atmosphere and in the granular layer. In the granular layer, radiative cooling balances the upward advection of energy. Cool, dense downflows drive the hot, less dense upflows by buoyancy, such that the radiative cooling in the cooling layer can be viewed as a driver for the convection. Correct treatment of radiation is therefore crucial also in the photosphere in order to obtain correct granular and subsurface topology. This has been pointed out by e.g. Stein & Nordlund (1989) and Nordlund & Stein (1991).

The full electromagnetic spectrum covers a few million spectral lines and photons can interact with the corresponding atomic or ionic transitions. A statistical treatment has therefore been necessary to make it possible to simulate this part of the sun within realistic time limits. First, the wavelength dependent opacities are sorted into four group mean opacities according to their magnitudes, and a similar treatment is given for the rest of the quantities needed for the radiative transfer problem (Skartlien 1998). Second, at each timestep I iterate the radiation field in each group, which can be visualized as propagating the photons from scattering event to scattering event until they finally get absorbed. This process is accelerated by using a new method given in Skartlien (1998).

Equations of mass, momentum, and energy are solved, and the exact form of the equations can be found in Nordlund & Stein (1990), along with de-

tails regarding the upper and lower boundary conditions, treatment of viscosity, and explicit predictor-corrector time stepping. Artificial diffusion is added to all quantities that are advanced in time to ensure stability. Spatial derivatives are taken with splines as basis functions. Horizontal boundaries are treated as periodic, such that outgoing flow at one side comes in at the opposite side. This is the case for both photons and fluid. Upper and lower boundaries are transmitting. The massflow out of the lower boundary is compensated by an equal inflow of mass to keep the total mass of the model constant. The entropy is given at the lower boundary. A constant upward energy flux is transported from the lower to the upper boundary. Outgoing shock waves at the upper boundary are almost fully transmitted.

The lower boundary is reflective for sound waves, such that a resonant cavity is formed with the photosphere as the upper reflective boundary for long period waves that are evanescent in the photosphere. The depth of the model convection zone is chosen such that the period of the radial mode (in the height direction) is 5 minutes, as to mimic 5 min. solar p-modes, and their influence on atmospheric dynamics.

The equation of state and quantities used for radiation (Skartlien 1998) are stored in a table that is calculated using packages of ionization equilibria and absorption coefficients (Gustafsson 1973), and opacity distribution functions (Gustafsson et al. 1975). Ionization and excitation of hydrogen and other abundant atoms, and formation of H_2 molecules, is included in the equation of state. At each timestep, I look up the needed quantities from the table by using the new density and internal energy as independent variables.

I use two low resolution runs with $32 \times 32 \times 100$ gridpoints. One run has the lowest possible viscosity without generating ripples, and one has higher viscosity to damp out long lived wave interference. These have a duration of 62 and 101 minutes of solar time respectively. One high resolution, low viscosity run with $64 \times 64 \times 100$ gridpoints is used as a validity check on the low resolution runs.

The processing time is large due to radiation iterations, and the hydrodynamic calculations take a negligible time in comparison. The high resolution run uses 10 000 CPU hours on 1 hour of solar time, on a SGI CRAY ORIGIN 2000 parallel computer. The radiation part of the code is run in parallel, with simultaneous computations for the radiation along different rays through the domain. The CPU time scales approximately linearly with the number of spatial gridpoints and angle points (rays) used for the radiation.

2.2 Diagnostics

Background physics. The main goal of this paper is to study why and how waves in the convectively stable atmospheric layers are excited when granules disappear. Vertical wave components immediately above disappearing granules have larger velocity amplitudes than the secondary horizontal wave components at the same height. I will therefore concentrate on a description of the excitation sources of vertical wave components by studying the equation of motion, and sources of pressure, pressure gradient and density changes. The equations that are discussed here are used in Sect. 3.2.

The vertical component of the equation of motion is:

$$\rho \frac{Du_z}{Dt} = -\frac{\partial P}{\partial z} - \rho g + \eta_z = F_z, \quad (1)$$

where ρ is density, D/Dt the Lagrangian time derivative (derivative following a fluid parcel), u_z vertical velocity, P pressure and g the acceleration of gravity which is taken as constant in our domain. F_z is net vertical force per volume unit. η_z is viscous force. The treatment of viscosity can be found in Nordlund & Stein (1990).

Pressure changes are due to adiabatic changes (density fluctuations) and non-adiabatic changes (entropy fluctuations). This is given by the first law of thermodynamics in the form:

$$dP = c^2 d\rho + (\Gamma_3 - 1)\rho T ds, \quad (2)$$

where $c^2 = \Gamma_1 P / \rho$ is the squared sound velocity with $\Gamma_1 = (\partial \ln P / \partial \ln \rho)_s$ at constant specific entropy s . $\Gamma_3 = (\partial P / \partial e / \rho)_\rho$ at constant density ρ . e is specific internal energy. T is temperature.

Usually, the differentials are substituted by the Lagrangian time derivatives (D/Dt), giving the changes in a parcel of gas following its motion. Here, I am instead interested in the time derivative of the pressure on a fixed point in space (Eulerian time derivative). Eq. (2) is also satisfied by a substitution with the Eulerian time derivatives³ (even if that means changes due to advection of flow past the point of measurement):

$$\frac{\partial P}{\partial t} = c^2 \frac{\partial \rho}{\partial t} + (\Gamma_3 - 1)\rho T \frac{\partial s}{\partial t}. \quad (3)$$

The time derivative of specific entropy is related to radiative and viscous heat sources and advection of specific entropy by:

$$\frac{\partial s}{\partial t} = \frac{1}{T}(q_{\text{rad}} + q_{\text{visc}}) - \vec{u} \cdot \nabla s. \quad (4)$$

³Similarly, the first law is also satisfied by spatial derivatives:

$$\nabla P = c^2 \nabla \rho + (\Gamma_3 - 1)\rho T \nabla s.$$

Using the continuity equation, and the last relation, we can write the time derivative of the pressure as:

$$\begin{aligned} \frac{\partial P}{\partial t} &= -c^2 \nabla \cdot (\rho \vec{u}) \\ &+ (\Gamma_3 - 1)\rho \{q_{\text{rad}} + q_{\text{visc}} - \vec{u} \cdot T \nabla s\}. \end{aligned} \quad (5)$$

The first term on the right hand side is the (adiabatic) contribution from mass flux divergence (the pressure decreases for a net mass outflow from a spatial point) and the second term is the (non-adiabatic) contribution arising from an imbalance between heating/cooling at a certain point and the advection of heat to that point (the pressure increases for a net gain of heat). The adiabatic and non-adiabatic contributions to the pressure are found by suitable time integrals over Eq. (3) or Eq. (5) in the following. This is useful for the purpose of discussing both horizontal and vertical flow.

The rate of change of the vertical pressure gradient also has contributions from one adiabatic and one non-adiabatic term. These are obtained by spatial derivation of Eq. (3):

$$\frac{\partial}{\partial t} \frac{\partial P}{\partial z} = \frac{\partial}{\partial z} \left\{ c^2 \frac{\partial \rho}{\partial t} \right\} + \frac{\partial}{\partial z} \left\{ (\Gamma_3 - 1)\rho T \frac{\partial s}{\partial t} \right\}. \quad (6)$$

I also discuss adiabatic and non-adiabatic contributions to the pressure gradient by time integration of Eq. (6).

Contributions from horizontal and vertical mass flux divergence control the density changes (via the continuity equation) and therefore also the adiabatic pressure and pressure gradient fluctuations. The mass flux divergence $\nabla \cdot (\rho \vec{u})$ is split into horizontal and vertical contributions:

$$-\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \vec{u}) = \nabla_{\text{h}} \cdot (\rho \vec{u}_{\text{h}}) + \frac{\partial}{\partial z} (\rho u_z), \quad (7)$$

where ∇_{h} is the divergence operator acting on the horizontal components \vec{u}_{h} of the velocity vector.

Reduction to the Klein-Gordon equation. Wave transients have been discussed using analytic solutions of the forced Klein-Gordon equation (Lamb 1908 and e.g., Rossi et al. 1992). This equation describes vertical, gravity modified, linear waves, and can therefore at best be used in atmospheric layers to describe small amplitude vertical waves (propagating or evanescent) if this is the dominating wave component. I compare simulation results with analytical predictions in later sections. The most important prediction from linear theory is the 3 min. eigenperiod of the solar atmosphere, and the decay rate of the transient oscillations produced by forcing of finite time duration. Even if quite restrictive assumptions are used to arrive at the Klein Gordon equation,

it predicts qualitatively correct results for the period and decay rate of the vertical wave components in lower atmospheric layers. This simplified description is however not adequate to explain the wave source, which involves nonlinear flow and horizontal flow contributions as shown in Sect. 3.2. For completeness, I derive the Klein-Gordon equation here and thereafter I discuss the predicted decay rates of the transient for different types of forcing.

Even if waves and convective motions are present we have small deviations from local hydrostatic equilibrium, which yields:

$$\frac{1}{\rho} \frac{\partial P}{\partial z} \simeq -g. \quad (8)$$

By combining the Eulerian time derivative of Eq. (1) (after dividing it by density), the thermodynamic relation in Eq. (5), the continuity equation, the approximation in Eq. (8), and neglecting the viscous force, I get:

$$\frac{\partial}{\partial t} \left(\frac{F_z}{\rho} \right) \simeq \frac{1}{\rho} \frac{\partial}{\partial z} \left\{ c^2 \nabla \cdot (\rho \vec{u}) - (\Gamma_3 - 1) \rho T \frac{\partial s}{\partial t} \right\} + \frac{g}{\rho} \nabla \cdot (\rho \vec{u}). \quad (9)$$

This equation governs the rate of change of the net vertical force per mass unit (at a fixed point), due to adiabatic and non-adiabatic changes in the pressure gradient (first term on the right hand side), and due to density changes in the gravity term (second term).

The Klein-Gordon equation results from a linearization of this equation, and by neglecting horizontal flow. If we assume isentropic (no entropy gradients), isothermal (constant sound velocity) and plane parallel conditions (zero horizontal mass flux divergence), and linear perturbations in u_z (negligible advection term in the acceleration), then Eq. (9) reduces⁴ to:

$$\frac{\partial^2 u_z}{\partial t^2} = \frac{c^2}{\rho} \frac{\partial^2}{\partial z^2} (\rho u_z) + \frac{g}{\rho} \frac{\partial}{\partial z} (\rho u_z). \quad (10)$$

We further assume that the gravity term can be neglected, and we keep only the first term, which is the pressure gradient term. Gravity comes only indirectly through the stratification of the density which is treated as a background density in hydrostatic equilibrium. Then we neglect $u_z \partial \rho / \partial z$ since the perturbation u_z is small in the linear limit. These simplifications lead finally to:

$$\frac{\partial^2 u_z}{\partial t^2} - c^2 \frac{\partial^2 u_z}{\partial z^2} = -\frac{c^2}{H} \frac{\partial u_z}{\partial z}, \quad (11)$$

⁴At this stage we could treat the linearized expression as a wave operator, and terms involving horizontal mass flux divergence and non-adiabatic and non-linear effects as wave sources. The Klein-Gordon equation is a further simplification of this wave operator.

where H is the density scale height. For infinite scale height (zero gravity) this becomes the ordinary wave equation. It can then be recast into the non-dimensional Klein-Gordon equation with the transformation $v = u_z e^{-z/2H}/c$:

$$\frac{\partial^2 v}{\partial \tau^2} - \frac{\partial^2 v}{\partial \zeta^2} = -v, \quad (12)$$

where $\tau = t/t_0$, $t_0 = 2H/c$, $\zeta = z/2H$. In the large wavelength limit, $H \ll \lambda$. We can then neglect the second derivative, and we see that the solution of the Klein-Gordon equation in this case gives a harmonic oscillation at frequency $\omega = 1$ which corresponds to dimensional frequency $\omega_a = c/2H$. This is the atmospheric eigenfrequency, or acoustic cutoff frequency. Inserting solar values, $H = 150$ km and $c = 10$ km s⁻¹, we get an oscillation period of about 3 min.

A driven solution (adding an inhomogeneous term) with a sinusoidal source beginning at $\tau = 0$ gives an evanescent wave for a driving frequency $\omega < 1$ and propagating waves (both upwards and downwards) for a driving frequency $\omega > 1$. Due to the sudden onset of the source, there will in both cases be a decaying oscillation at $\omega = 1$ with amplitude proportional to $t^{-3/2}$ (asymptotically) and a full spectrum of initial transients that decay faster in time. The amplitude would increase to infinity for driving at the resonance frequency $\omega = 1$, since viscous forces are neglected. A delta function as a source produces a flat spectrum of both evanescent and propagating components which also decay in time. The eigenoscillation at $\omega = 1$ decays in this case asymptotically as $t^{-1/2}$. The decay time of the oscillations would be shortened if viscous forces were added.

It should be noted that the wave amplitude for u_z for propagating waves increases exponentially with height (v is constant in this case), such that this linear theory predicts its own breakdown. At large enough heights, we expect wave steepening and shock formation due to the advection term in the equation of motion.

Envelope characterization of transients. It is also of interest to determine the transient response of the model atmosphere to the acoustic events in the simulation data. In general, a continuous spectrum of oscillation periods will be excited by a forcing of finite duration. I study the amplitude variations in time for different oscillation periods, and determine their relative amplitudes. I use two frequency bands, one centered at 5 min. period (p-modes in the convection zone of the simulation) and the other at 3 min. period (atmospheric eigen-period). In this way, I can determine a possible redistribution of energy between

the frequency bands as a function of time, or whether p-modes or atmospheric oscillations are excited.

Since I consider wave packets with finite time duration (transients), I can either use a wavelet transform to display amplitude and phase as function of time and frequency, or, as I have done here, use the analytic signal, y_a that decomposes the measured signal y_r into an envelope (amplitude of the wave packet) $A(t)$ and a phase $\Phi(t)$ as a function only of time⁵ (see e.g. Bracewell 1986):

$$\begin{aligned} y_a &= A(t) e^{i\Phi(t)} & (13) \\ y_r &= \text{Re}[y_a] = A(t) \text{Re}[e^{i\Phi(t)}]. & (14) \end{aligned}$$

Here, $\text{Re}[\dots]$ means the real part. The measured signal is the real part of the analytic signal. The complex component of the analytic signal is introduced by the exponential term.

I calculate the analytic signal after bandpass filtering of the measured signal, and obtain the amplitude and phase within each of the two frequency bands. The analytic signal is easily obtained by Fourier transformation only. The details in calculating envelope and phase, and the importance of band limitation, is explained in the appendix.

Particle paths. An efficient and simple way to visualize atmospheric wave motion and convective fluid flow is by plotting particle paths of fluid elements. This will be used frequently in the following. A particle path is in general given by:

$$\vec{\xi}(t) = \int_0^t \vec{u}(\vec{\xi}(t'), t') dt' + \vec{\xi}(0), \quad (15)$$

where $\vec{\xi}(t)$ is the vector pointing at the position of a particle that is located at $\vec{\xi}(0)$ initially. \vec{u} is the velocity vector that depends on position and time. If we take the double time derivative of one of the components of $\vec{\xi}$, say the vertical component ξ_z , we get the Lagrangian acceleration in that direction:

$$\frac{d^2 \xi_z}{dt^2} = \frac{Du_z}{Dt} = \frac{\partial u_z}{\partial t} + u_z \frac{\partial u_z}{\partial z} + \vec{u}_h \cdot \nabla_h u_z = a_z, \quad (16)$$

where \vec{u}_h is the horizontal component of the velocity vector, and ∇_h the horizontal components of the gradient operator. u_z is the vertical velocity, and a_z the vertical acceleration made up of all the forces in the vertical direction. I have separated the horizontal advection of u_z from the pure vertical term $u_z \partial u_z / \partial z$ to emphasize the influence of horizontal motion on vertical acceleration.

⁵The analytical signal can be obtained by integration of the wavelet transform along the frequency axis. This produces the instantaneous amplitude and phase.

I analyze simulation data from a vertical column (fixed horizontal position) centered in the collapsing granule. Horizontal velocities \vec{u}_h are small compared to the vertical velocity u_z in the initial phases of the collapse. The particles found in the column will then stick to it (move up and down along it), as long as \vec{u}_h is negligible. ξ_z can then be integrated by taking u_z from this column only. A plot of the ξ_z curve in the time-height plane will in this case show a curvature that corresponds to:

$$\frac{d^2 \xi_z}{dt^2} = \frac{\partial u_z}{\partial t} + u_z \frac{\partial u_z}{\partial z} = a_z^*, \quad (17)$$

where a_z^* is the acceleration derived from vertical forces (per unit mass) in the column. It is this expression I use for visualization.

When horizontal velocities come into play, then $\vec{u}_h \cdot \nabla_h u_z$ is no longer zero, and the particle path derived from a_z^* in the column will not correspond to a true particle path. This will be the case when neighboring granular cells move into the column after the collapse, and the true particle paths will cross the column horizontally.

3 Results

I describe some signatures of the events in Sect. 3.1. First, I show quantities that can also be measured in solar observations (observable signatures). These signatures are the continuum intensity and vertical velocity in the evolving granular pattern during an acoustic event, in addition to the response in vertical velocity in the overlying atmosphere. Then I show *unobservable* signatures in the flow evolution above and below the cooling layer, and a correlation to meso-granular flow beneath the cooling layer. I turn to more physical descriptions of the wave excitation in Sect. 3.2.

3.1 Description of acoustic events

I have made two simulation series, which both show disappearing granules in the cooling layer and subsequent atmospheric oscillations. The first is with low viscosity that includes high frequency, short wavelength wave components, and shock waves in the chromosphere. The second series has higher viscosity, which serves to damp the short wavelength components. The advantage of this damping is that the atmospheric oscillations from the events become more distinct due to less short wavelength interference. It was therefore easier to unambiguously detect the cause-effect relationship between the collapsing granules and the atmospheric oscillation.

This high viscosity simulation does not include many shock structures, and the granules are quite smeared in comparison to what is observed. I therefore returned to the more realistic low viscosity simulation after I discovered that collapsing granules excited atmospheric oscillations. I will mainly discuss the low viscosity results in the following.

3.1.1 Observable signatures

Granular pattern and wave excitation. Correlations between changes in the granular pattern and excitation of atmospheric oscillations were found by comparing movies of horizontal slices of vertical velocity in the atmosphere at $z = 0.7$ Mm and in the granulation layer at $z = 0.0$ Mm. In general, I find that rapid topology changes in the granular cells correlate to the onset of horizontal and vertical wave components at $z = 0.7$ Mm. These events are disappearing small granules and fragmentation of larger granules.

The strongest correlations are found between single, vanishing small granules and horizontally localized, vertical velocity oscillations at $z = 0.7$ Mm. Horizontal wave components are also detectable in some of these events, and appear as horizontally expanding rings.

Small fragments of larger granules that disappear in intergranular lanes immediately after the splitting process, produce in some cases a response in atmospheric layers, but with a larger extension of the perturbations, and with weaker amplitudes than for the single cell events. Horizontal waves are also seen here in some cases but they appear as elongated structures with stronger amplitudes on one side of the event.

In the following, I discuss a time sequence that shows characteristic properties of single cell events in the low viscosity run. This is shown in Fig. 1, as horizontal slices of vertical velocity at $z = 0.7$ Mm and at $z = 0.0$ Mm, that evolves from the top to the bottom in the figure. The evolving granular pattern appears from the velocity images at $z = 0.0$ Mm. The smeared continuum intensity (granulation intensity) is also shown, as to mimic observations. Finally, we see the vertical component of the kinetic energy flux, $\rho u^2 u_z / 2$ and the vertical component of the wave energy flux,⁶ Pu_z . These fluxes are both measured at

⁶We note that the interpretation of $P\bar{u}$ (or its vertical component Pu_z) as wave energy flux is only meaningful in the absence of convective flow fields, or when the convective flow field evolves on a much longer timescale in comparison to the period of the waves considered. This fact is often overlooked when acoustic fluxes with oscillating components of e.g. 5 min. or 3 min. are measured from observations. These periods are clearly not much shorter than the typical lifetime of granular cells. The oscillating velocity pattern seen at $z = 0.7$ Mm in the simulations can be attributed mainly to wave motion since the velocity from the convective overshoot is dramatically reduced above $z = 0.5$ Mm, and it is relatively small in comparison to

$z = 0.7$ Mm.

A few typical features can be summarized as follows. The event is generated by a small, isolated granule in the vertex of intergranular lanes. The small granule shrinks horizontally, upflow reverses to downflow, and an extended downdraft forms in the vertex. This is seen at the coordinate (4.0, 0.5) Mm in the figure (at the border between the images in the upper right corner). The upflow in the granule reverses to downflow at $t = 0.0$ min. In the following, this flow reversal defines the time when the granule collapses or disappears.

A dark spot in the smeared continuum intensity starts to develop around this vertex after the granule has disappeared. This is seen in the middle column in the figure. The white intensity contour marks the 90% level relative to the average intensity in the time sequence, and the dark spot has intensity levels below that.

The overlying atmosphere at $z = 0.7$ Mm above the vertex, responds with a relatively small downward velocity as seen in the figure at $t = 0.5$ min. A larger upward velocity pulse follows at $t = 2.5$ min. A succeeding smaller downward velocity is found at $t = 4.5$ min. Then a decaying oscillation follows at this position. The horizontal size of the oscillating spot is initially about 1.0 Mm, but expands when a circular shaped wave propagates outward in the horizontal direction. This is seen at $t = 5.5$ min in the figure.

The vertical wave energy flux Pu_z at $z = 0.7$ Mm shows upward and downward pulses in phase with the velocity, and about equal duration of upward and downward fluxes. The initial upward flux has larger magnitude than the following downward pulse. This means that energy is transported upwards at this height. The vertical kinetic energy flux $\rho u^2 u_z / 2$ shows a similar behavior.

Observations of vertical wave energy flux in the photosphere (Rimmele et al. 1995) show upward directed wave energy flux accompanied by a darkening of intergranular lanes. The perturbation in wave energy flux is measured at the same horizontal position as the darkened intergranular lane. This is consistent with the upward directed acoustic wave energy flux (Pu_z -flux) as discussed above, and the emergence of the dark spot in the continuum intensity above the intergranular lanes. Furthermore, Espagnet et al. (1996) observed darkening in the granulation with an accompanying transient intensity oscillation that also supports the simulation results.

I find additional low energy fluctuations superposed on the images that are not directly related to the vanishing granule events, but they may be due

the oscillation amplitudes caused by the events.

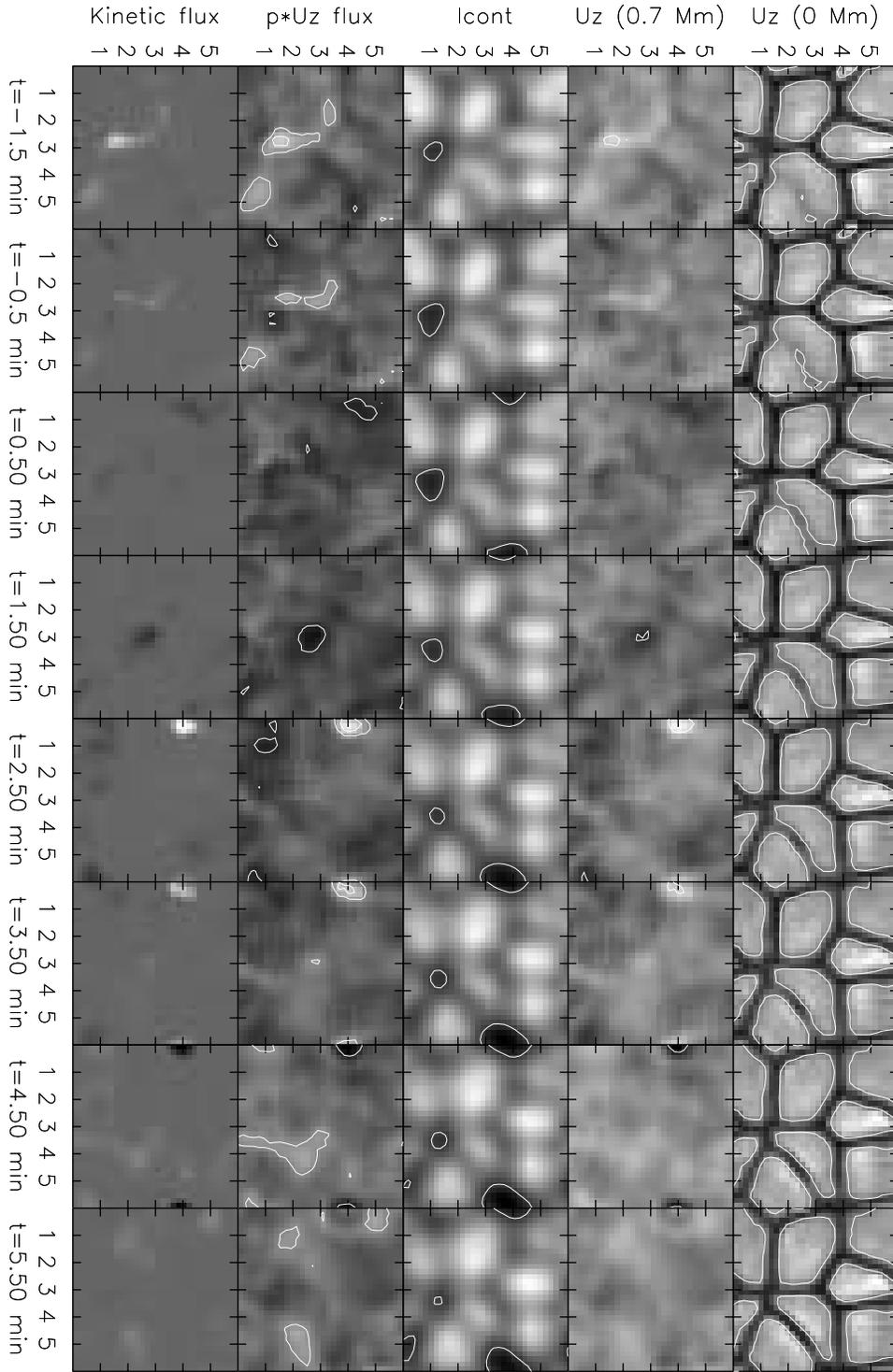


Figure 1: *Observers view*. Each column shows the evolution of different quantities in an acoustic event, and each image is a horizontal slice that covers the full horizontal extent of 6×6 Mm. The time series starts at the top and continues downward. The quantities are, starting at the rightmost column: $U_z(0 \text{ Mm})$: Vertical velocity in the cooling layer at $z=0.0$ Mm. The contour shows zero velocity. Upward velocity corresponds to lighter shades of grey. A small granule disappears at $t=0.0$ min, at the coordinate $(4.0, 0.5)$ Mm (at the border between the images in the upper right corner). $U_z(0.7 \text{ Mm})$: Vertical velocity at 0.7 Mm above the cooling layer. The contours show ± 0.6 times the maximum velocity in the series. After the granule disappears, we see an upward velocity pulse at $t=2.5$ min, also at the position $(4.0, 0.5)$ Mm. A downward pulse follows, which is seen at $t=4.5$ min. A circular wave pattern can be seen at $t=5.5$ min. **lcont**: Smearred continuum/granulation intensity. The contour marks 0.9 times the average intensity. An apparent dark spot grows after the granule has disappeared. **Pu_z -flux**: (0.7 Mm) The contours show the levels ± 0.6 and ± 0.3 times the maximum flux in the series. The upward flux is larger than the downward flux. **Kinetic flux**: Vertical kinetic energy flux $(\frac{1}{2}\rho u^2)u_z$ at 0.7 Mm.

to wave motions originating from previous events, or by other granular motions such as splitting cells. A splitting granule is also seen in Fig. 1, but that does not produce these highly localized pulses with high energy flux density (per horizontal area).

Correlations in vertical velocity. Simultaneous velocity curves from the granular layer at $z = 0.0$ Mm and the atmosphere above at $z = 0.7$ Mm are shown in Fig. 2 for 6 different events found during 65 min of solar time in the low viscosity simulation. These velocities are taken from horizontal positions defined by the centers of the vanishing granules at $z = 0.0$ Mm. The time of granular velocity reversal is at $t = 0.0$ min, as in the previous paragraph.

The figure shows, for the vertical velocity at $z = 0.7$ Mm, a time delay spread (after $t = 0.0$ min) from 1 to 3 min for the maximum upward velocity pulse. The pulse magnitudes are $1 - 2 \text{ km s}^{-1}$. There is also a weak tendency of larger maximum upward velocity pulse for larger time delay. After the upward pulse, a decaying oscillation starts.

I have also calculated an average response of the vertical velocity, which is also shown in Fig. 2. Here, I align the velocity sequences such that maximum upward velocities coincide before I average all curves. This is necessary since we have different time delays for the maximum. On the average, the first upward velocity phase lasts for 2 minutes, with a maximum amplitude of 1.25 km s^{-1} . There is a preceding downward velocity of -0.5 km s^{-1} and a larger succeeding downward velocity of -0.75 km s^{-1} .

A remarkable fact is that, on the average, the atmosphere responds to the newly formed downdrafts with a delayed upward velocity pulse that is larger than the initial downward velocity. The cause of this behavior will be discussed later in Sect. 3.2.

The granular velocity for the events evolves almost linearly in time around $t = 0.0$ min., and we note that this corresponds to ballistic (parabolic) motion. Most of the velocity jump is covered within a time span of 2 min. The steep decline of the granular velocities is evidently connected to the wave transient at $z = 0.7$ Mm, as seen in the figure. We also note that the upflow velocity in the granule is *increasing* before the collapse.

The different strengths of the events (as measured by the velocity magnitude at $z = 0.7$ Mm) are not obviously linked to the timescale or magnitude of the granular velocity jump. Atmospheric oscillations may exist prior to the excitation of the events. In these cases, the interference from atmospheric oscillations modulates the velocity signal at $z = 0.7$ Mm, which makes such a correlation difficult. The amplitude of the atmospheric velocity pulse is in these cases de-

pendent on whether the excitation comes in phase or out of phase with the preceding atmospheric oscillation. Despite the modulation by interference, the average velocity curves in the lower right panel in Fig. 2, should give a reliable picture of the atmospheric velocity response.

As indicated in Fig. 2, only 3 events of 6 show clear (detectable) signatures of horizontal wave propagation as indicated by the labels **ws** and **wa**. The single cell events, (A) and (B) (as indicated in the figure), with additional horizontal wave propagation, will be analyzed in detail. Event (B) is picked to study the physics behind the events in Sect. 3.2.

The same type of plot for the high viscosity events is given in Fig. 3. Here I show 6 different events, found in a timespan of 101 min. in solar time. The transients are in this case very clean due to less interference with other waves. It is clear that only the single cell events (**sc**) give a significant transient vertical oscillation. Even if splitting cell events (**spc**) do not show vertical oscillations, they may show asymmetric horizontal wave fronts as indicated by one example in each of the high and low viscosity runs. Note also that the exploding granule event, (**exc**), shows a relatively slow change of the granule velocity, and therefore no atmospheric response.

An estimate for the solar event rate. A useful signature of the events is their occurrence rate that can be compared to observations. This is interesting for both photospheric velocity measurements and the occurrence rate of the chromospheric bright grains. For the low viscosity run, and counting all events (7 in total), I find $3 \times 10^{-3} \text{ events min}^{-1} \text{ Mm}^{-2}$ which corresponds to approximately $2 \times 10^4 \text{ events min}^{-1}$ for the whole surface of the sun. The lower event rate $2 \times 10^{-3} \text{ events min}^{-1} \text{ Mm}^{-2}$ for the high viscosity run is expected, since the granular motions are less vigorous and there are fewer smaller cells. For the real sun, the viscous effects are much smaller than in the low-viscosity simulation, such that the estimate above can only provide a lower bound on the event rate.

Excitation of 3 min versus 5 min oscillations. The vertical velocity at $z = 0.7$ Mm is bandpass-filtered using two frequency bands. One filter is centered at 5 min, which is the period of the radial p-mode oscillation with the largest amplitude. These waves are propagating in the convection zone, but form a standing wave in the cavity between the lower boundary of the simulation and the cooling layer. This p-mode is evanescent in the atmosphere since the period is longer than the acoustic cut-off period. The other bandpass filter is centered at 3 min, which

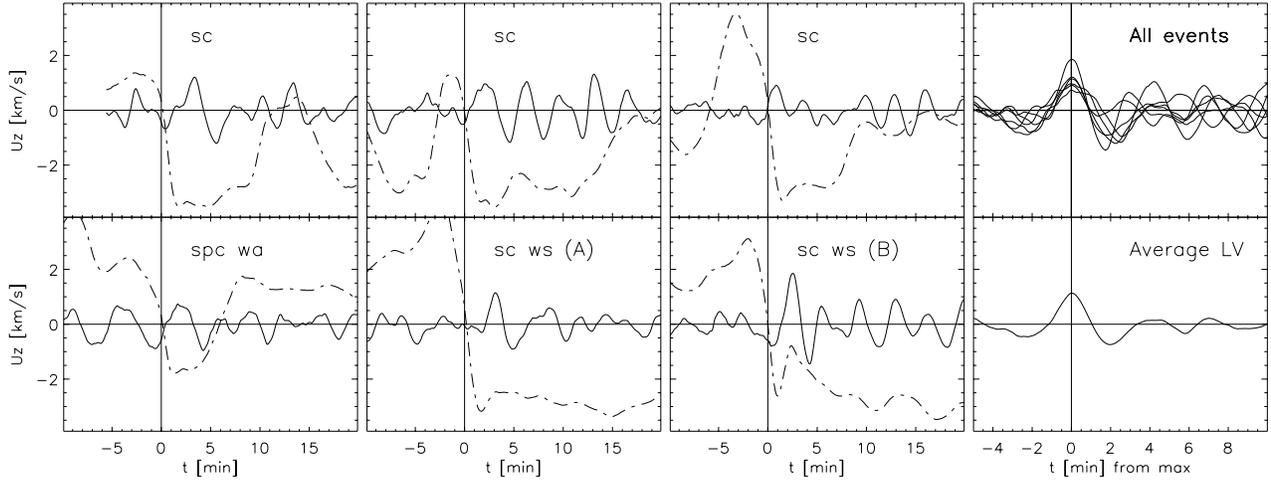


Figure 2: *Six events in the low viscosity run (65 min. solar time).* The solid line is the vertical velocity at $z = 0.7$ Mm, and the dash-dotted line the vertical velocity in the granule at $z = 0.0$ Mm. Each velocity signal is measured at the horizontal position centered above the vanishing granule. At $t = 0.0$ min, the vertical velocity in the granule reverses from upward (positive) to downward (negative) velocity. The two rightmost panels show all velocity curves for the vertical velocity at $z = 0.7$ Mm, and the corresponding average. The time axis is in this case centered on maximum upward velocity for each individual event. The codes in each panel for the individual events are as follows: **sc**: single granule that vanishes. **spc**: splitting granule, which means that a fragment separates from a larger granule, and vanishes in a downdraft immediately afterwards. **ws**: circular shaped, horizontally propagating wavefront at $z = 0.7$ Mm. **wa**: asymmetric wavefronts, larger on one side of the event.

The time delays for the upward velocity pulse in the atmosphere for the events, lie in the interval [1, 3] min, with pulse magnitudes in the interval [1, 2] km s^{-1} . On the average (lower right panel for all events), the upward pulse lasts for about 2 min, with an amplitude of 1.25 km s^{-1} . We see a preceding downward velocity of -0.5 km s^{-1} and a larger succeeding downward velocity of -0.75 km s^{-1} . The granular velocity evolves almost linearly in time around $t = 0.0$ min. when the timescale is shortest. Most of the velocity jump is covered within a timespan of 2 min.

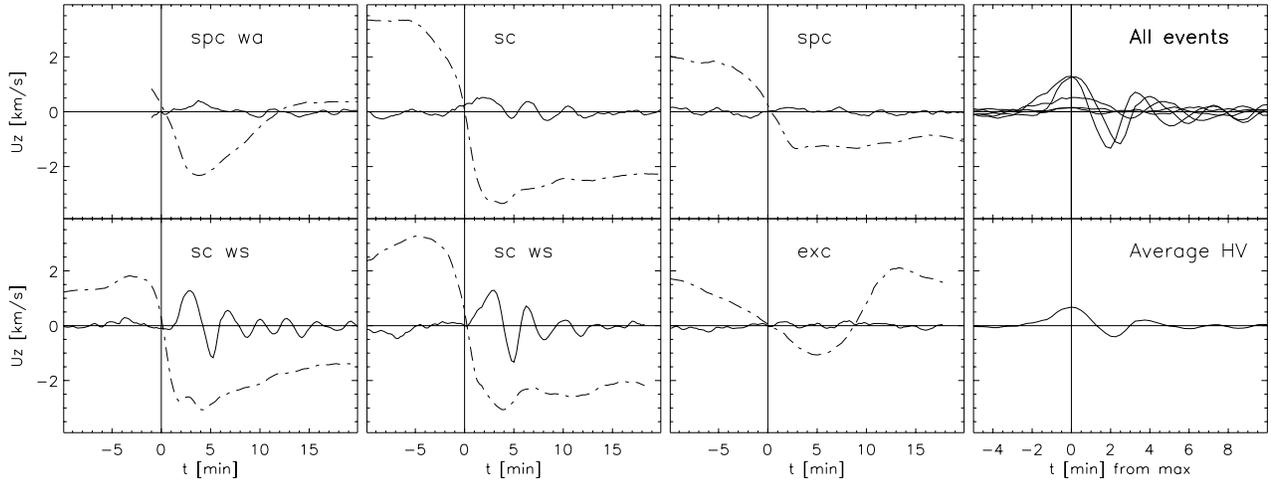


Figure 3: *Six events in the high viscosity run (101 min. solar time).* The meaning of the velocity curves is the same as in Fig. 2. The codes in each panel for the individual events are the same as in Fig. 2, with the addition of **exc**, which means that a large granule collapses in the middle, and splits in several fragments (exploding granule).

The time delays for the upward velocity pulse in the atmosphere for the **sc** events, lie in the interval [2, 3.5] min, with pulse magnitudes in the interval [0.5, 1.5] km s^{-1} . The velocity amplitudes are smaller and the delays are larger than for the low viscosity case. On the average (lower right panel for all events), the upward pulse lasts for about 3 min (which is longer than for the low viscosity run). As with the low viscosity case, the granular velocity evolves almost linearly in time around $t = 0.0$ min when the timescale is shortest, and most of the velocity jump is covered within a timespan of 2 min.

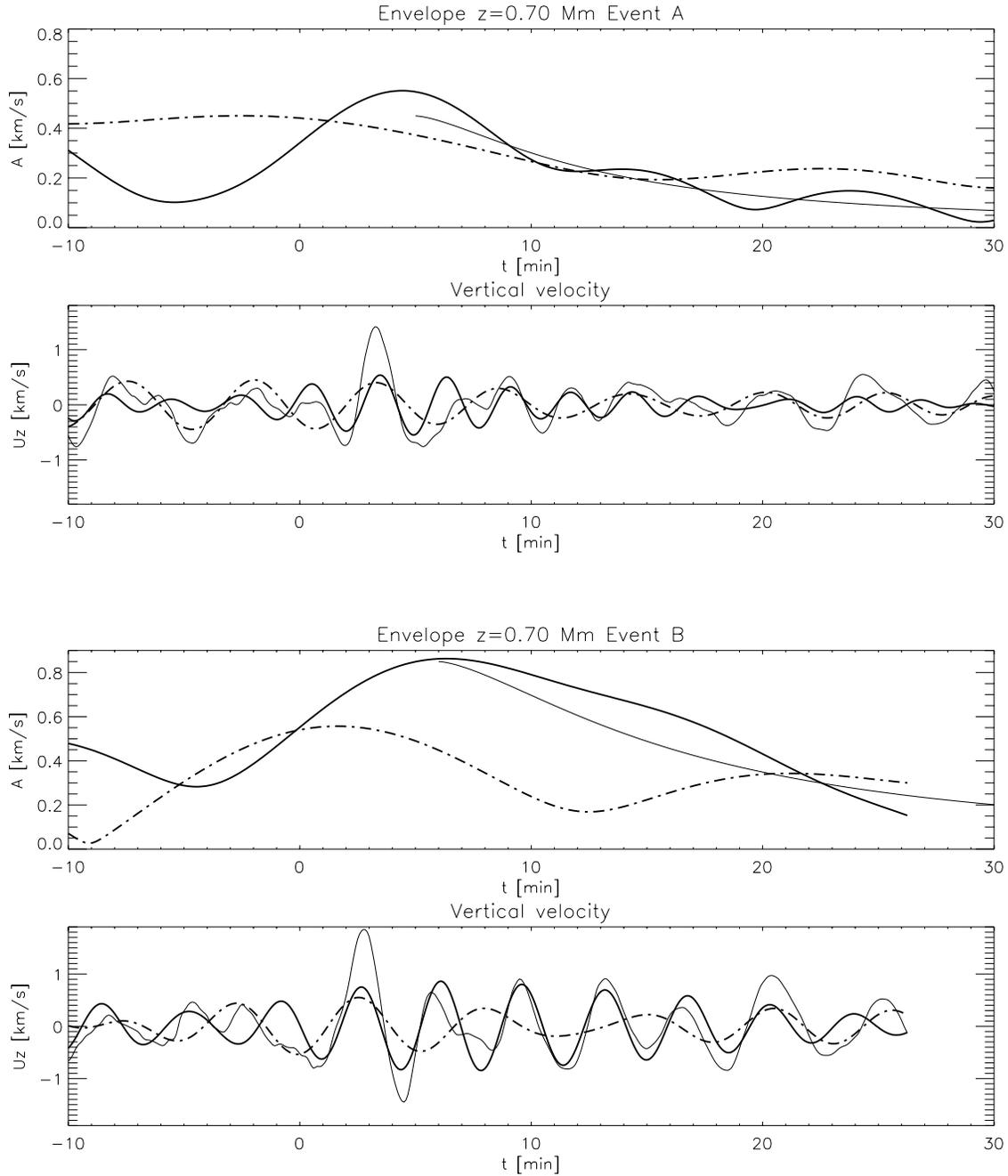


Figure 4: *Envelope decomposition of events A and B.* The upper panel for each event shows the envelope of the vertical velocity oscillation in the 5 min (dash-dotted line) and 3 min (solid line) frequency bands. The velocity signal is taken at $z = 0.7$ Mm, above the vanishing granule. The time is zero when the granule vanishes at $z = 0.0$ Mm. The thin solid line indicates the best fit to a $t^{-3/2}$ decay of the envelope. This is the analytical solution of the asymptotic decay of the 3 min. eigen-period, as described by the Klein-Gordon equation, when this is driven by sinusoidal forcing starting at $t=0.0$ min.

It is noted that the amplitude of the envelope for the 3 min band increases at the onset of the events, reaches a maximum at about $t=5.0$ min, and then decays approximately as predicted from linear theory. The envelope for the 5 min band is damped rather than amplified after the onset of the acoustic events. Hence, the events are mainly exciting oscillations in the 3 min band. The lower panel for each event shows the bandpass filtered velocities that correspond to the envelopes in the upper panel. The envelopes are enclosing these “wave packets”. The original velocity signal is shown in addition, by the thin solid line. It is also noted that the maximum velocity amplitude (at about $t=3.0$ min) precedes the peak in the envelope in the 3 min band.

is the eigen-period of the atmosphere. The widths of the filters are chosen such that they cover most of the power in the neighborhood of these periods in the Fourier domain.

I have thereafter computed the time dependent envelope for each of the two bandpass filtered signals as described in Sect. 2.2. One can then determine the relative contribution from the wave excitation (caused by the event) to the 5 min p-mode and the 3 min atmospheric mode. This procedure is applied to events (A) and (B). The envelopes and the corresponding bandpass filtered signals are plotted in Fig. 4.

We see that both events have dominating 3 min components after the initialization of the transient at $t = 0.0$ min (upper panels). The envelope of the 3 min band grows from $t = 0.0$ min (when the cell collapses) and reaches the peak at $t = 5.0$ min, after which it decays. In contrast, the envelope of the 5 min band decays immediately after $t = 0.0$ min. The 5 min envelope magnitudes are also insensitive to the velocity amplitude of the events at $t = 0.0$ min (thin line in the lower panels), as opposed to the 3 min envelopes.

This indicates an excitation of the eigen-oscillation of the atmosphere by the acoustic events, and that energy is fed into the 3 min oscillation rather than the 5 min oscillation.

The asymptotic $t^{-3/2}$ decay for the 3 min eigen-oscillation from linear and plane parallel theory (Klein-Gordon equation from Sect. 2.2) fits well for event (A) but not very well for event (B). Discrepancies from the smooth, analytical prediction can also here be due to interference from waves that are not excited by the events.

3.1.2 Unobservable signatures

I use only the low viscosity results in the following. First, I show that collapsing granules are predominantly located above subsurface downflows as deep as 1000 km below the cooling layer. Then I give a description of the convective motions that precede the cell collapse and the wave motions in the atmosphere that follow the collapse. I use event (B) as an example. Physical explanations of this flow are given later in Sect. 3.2.

Correlations with mesogranular flow. Figure 5 shows the convective flow below the surface, and its connection to the vanishing granules. Two horizontal planes from two different levels below the surface ($z = -0.5$ Mm and $z = -1.0$ Mm) are shown for each vanishing granule event. The white arrows show horizontal motion of the granules in the surface ($z = 0.0$ Mm) before the collapse. These arrows start where a fragment splits off a larger granule and forms a smaller granule (as it appears at the surface). The

arrows end where the granules collapse, as measured by zero vertical velocity at $z = 0.0$ Mm.

The intergranular lanes are still visible at $z = -0.5$ Mm. At $z = -1.0$ Mm, the downflows are more disconnected, but they still tend to form a network pattern delineating “mesogranular” upflow regions. The intergranular vertices at $z = -0.5$ Mm are approximately positioned above relatively high velocity downdraft regions at $z = -1.0$ Mm.

The sites where granules collapse at the surface also tend to be located approximately above these downdrafts at $z = -1.0$ Mm, and near intergranular vertices at $z = -0.5$ Mm. As seen by the arrow directions, there is also a tendency of horizontal advection of the granules towards these locations. This advection is a consequence of pressure inequalities in the neighborhood of the cell (large pressure in large cells), driving it in the direction of lowest pressure. This suggests that collapsing granules at the surface are separated by a typical length scale corresponding to the mesogranular scale of the subsurface convective flow.

There is not a clear correlation between horizontal velocity vectors in the subsurface layers as shown in the figure, and the advection direction of the granules at the surface. This is because the velocity field in the figure is taken from a snapshot at the time of cell collapse (that does not necessarily correspond to the advection direction), and because the surface and subsurface velocity fields differ.

Flow above and below the cooling layer. I have displayed a time sequence of the flow in a vertical plane that cuts through event (B) in Figs. 6 and 7. The contour for zero vertical velocity in Fig. 6 encloses the vanishing convective upflow as a vertical “tongue” at about $y = 4.0$ Mm in the upper left panel ($t = -2.0$ min). A downflow, originating from an intergranular lane, is deflected sideways into the deeper parts of the upflow at about $z = -1.0$ Mm.

The gray scaled images show the relative pressure deviation from the horizontally averaged background. As the granule and the associated upflow vanish, the pressure decreases locally in the convective overshoot zone between $z = 0.0$ Mm and $z = 0.5$ Mm. This is seen for $t = 0.0$ min and $t = 1.25$ min. This pressure deficiency sets up horizontal pressure gradients, that accelerate the flow horizontally towards the site of collapse, and horizontal mass flux convergence is initiated. The pressure increases ($t = 2.08$ min) due to this inflow (as discussed in Sect. 3.2), and reaches a maximum around 3 min. This is seen at $t = 2.91$ min in Fig. 7. Hence, a negative pressure perturbation occurs in the overshoot zone, when convective upflow reverses to downflow.

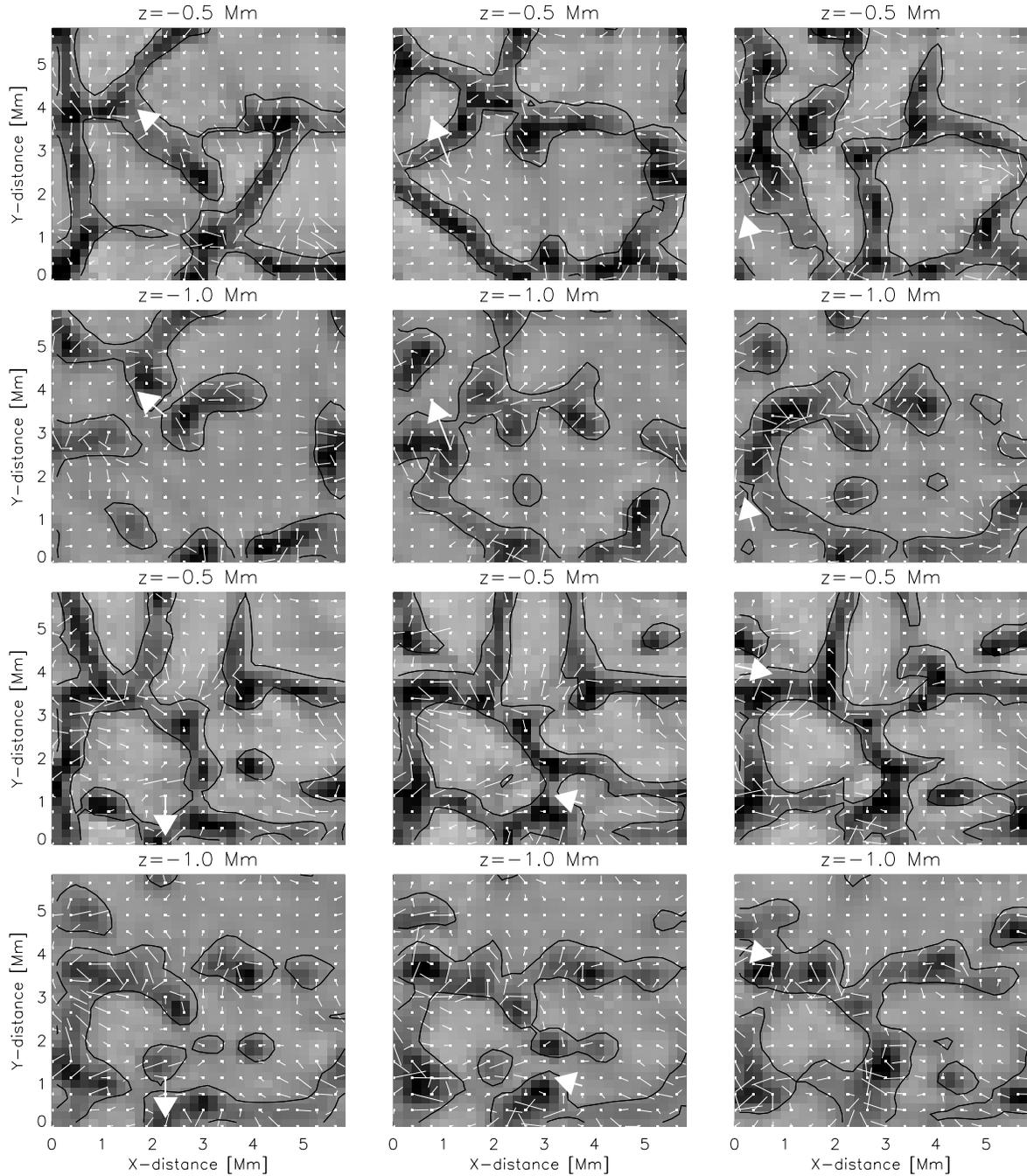


Figure 5: *Correlation to subsurface flow.* Shaded images: vertical velocity at 0.5 Mm (upper images) and 1.0 Mm (lower images) below the surface, with upflow as lighter shades. Corresponding images (two simultaneous images for each event) are vertically aligned. Black contours: zero vertical velocity. Velocity vectors for horizontal velocities are white “flags” pointing in the “down-wind” direction. White arrows show the motion of the central parts of small granular cells at the surface (0.0 Mm), from where the cell splits from a larger cell (starting point), to the point where they collapse (end points). The time and position of collapse is defined by the last measurement of upward velocity in the surface. All velocity images are taken at the time of cell collapse in the surface.

The horizontal position at the surface where granules collapse tends to be located above, or near subsurface downflow. In most cases, there is also a tendency of “attraction” to these locations, as seen by the arrows.

A stagnation point (upflow above and downflow below) has developed at about $y = 3.7$ Mm, $z = 0.2$ Mm and $t = 2.91$ min. The upward velocity at $z = 0.7$ Mm is now maximum, with a corresponding pressure maximum, as in a propagating acoustic wave. This locally enhanced pressure has produced horizontal expansion and horizontal wave components at $t = 4.16$ min, between $z = 0.5$ Mm and $z = 1.0$ Mm. As the positive pressure perturbation propagates horizontally outwards in both directions, the pressure at the site of the initial perturbation decreases. This is seen as the “flattened donut” centered at $y = 3.8$ Mm, $z = 0.8$ Mm. The horizontal propagation continues at $t = 5.41$ min and $t = 5.83$ min. The horizontal, longitudinal displacements are driven by horizontal pressure gradients, and the vertical transverse displacements by buoyancy. It is this wave that appears as a horizontally expanding circular ring in vertical velocity (see observers view in Fig. 1).

The density contours in the sequence display V-shaped, horizontally expanding phase-fronts that correspond to a downward directed wave vector. These fronts are seen by drawing a line through the peaks of adjacent density contours. The horizontal and vertical components of the wave vector, together with an oscillation period of about 3 min, locate this wave mode in the gravity wave domain. The period would have been considerably shorter than 3 min if these oscillations were acoustic with the same direction of the wave vector. The Brünt Väisälä period is slightly lower than the acoustic cutoff period of 3 min, hence these wave components are barely within the gravity wave domain. Furthermore, they do not survive for more than 2-3 oscillation periods in the vicinity of the event before they are washed out by interaction with flow and other wave components. The wave amplitude is in addition damped by the geometrical effect of expansion.

At $t = 5.41$ min, gas flows downwards to the negative pressure perturbation at $y = 3.8$ Mm and $z = 0.8$ Mm. The downward moving gas falls into the relatively slow moving gas at around $z = 0.8$ Mm, resulting in non-linear steepening of an upward propagating wavefront at $z = 1.2$ Mm and $t = 5.83$ min. This wavefront becomes supersonic at a later stage, and belongs to the family of gravity modified acoustic waves.

3.2 Physics of acoustic events

First, I discuss the vertical wave components that are excited in the atmosphere by using events (A) and (B). Thereafter I discuss the excitation of these waves by first considering convection dynamics that lead to the granule collapse (Sect. 3.2.1), and then by considering the resulting wave excitation in the con-

vective overshoot flow above the collapsing granule (Sect. 3.2.2).

Vertical wave components and acceleration.

In response to the reversal from upflow to downflow in the collapsing granule, the fluid in the atmosphere is initially accelerated downwards. This is followed by an upward acceleration, and a decaying oscillation follows. Figs. 8 and 9 show the vertical acceleration (grey scale image) and the associated particle paths (white curves) along the column centered on the events (B) and (A) respectively. The particle paths show vertical positions of fluid parcels as discussed in Sect.2.2. The black contours mark locations of zero vertical acceleration. They enclose a region of positive (upward) or negative (downward) acceleration. If the particle paths show positive curvature (positive second time derivative) the acceleration is positive. $t = 0.0$ min when upflow reverses to downflow in the granule.

The fluid in the atmosphere (above $z = 0.0$ Mm) moves down initially at $t = 0.0$ min in response to the downward acceleration. The downward velocity reverses to upward around $z = 0.7$ Mm for $t = 2.0$ min (as seen by the slopes of the particle paths). The upward acceleration that is associated with the wave excitation is seen as the slanted light grey region around $t = 2.0$ min which extends down to 100 km above the surface. Note that this is seen for both events.

Upward propagating waves (slanted contours) are evidently present in the wave train that follows. Approximately one half wavelength, or even less, fits within the interval [0.3,1.5] Mm up to the upper boundary. The 3 min period is present in both events, as also shown previously. These long wavelengths indicate strongly gravity modified acoustic waves, as is expected for oscillation periods close to the cutoff period. The oscillations have an evanescent character below $z = 0.5$ Mm, and are visible in the figures down to $z = 0.3$ Mm (seen as sign changes of the acceleration). The convective overshoot flow is significant at this level, and adds an almost constant downward acceleration to the small oscillatory component.

For event (B), we see relatively small amplitude oscillations in atmospheric layers with a dominating period of about 2.0 min before the transient is excited. It appears from the figure that the upward acceleration of this oscillation is approximately in phase with the upward acceleration from the wave excitation. This serves to amplify the displacements and vertical velocity of the wave transient. This effect is not present, or is much weaker for event (A), for which the the atmospheric oscillation prior to the excitation is less coherent and has smaller amplitude. The atmospheric response for event (B) is, in spite of the extra

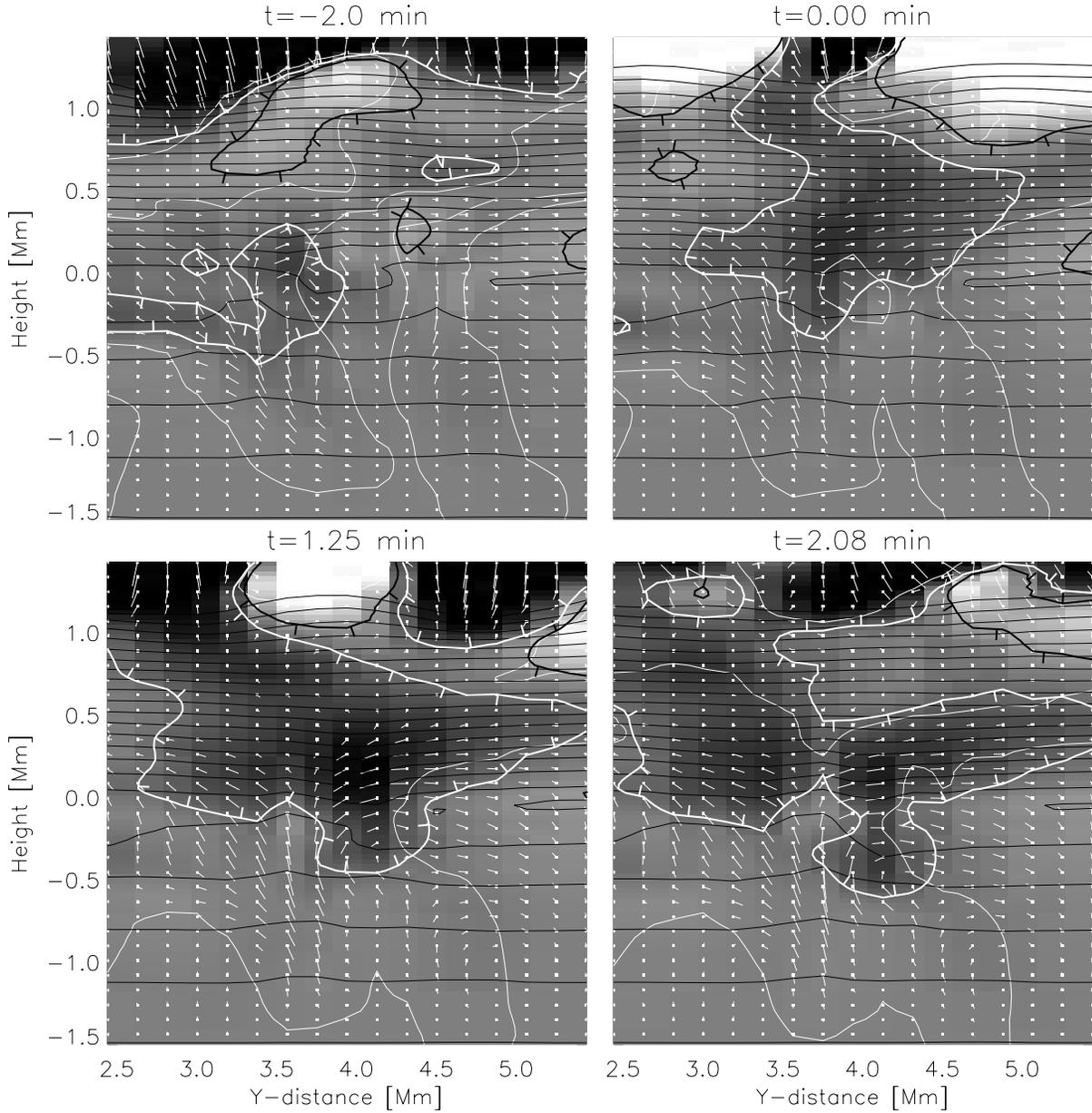


Figure 6: *Cell collapse and photospheric pressure perturbation, event (B)*. Shaded picture: relative pressure perturbation from the mean pressure. Black and white contours: +10% and -10% pressure perturbations respectively. The ticks on these contours point in the downhill direction (away from the positive pressure gradient perturbation). Black contours are density-contours. White contours: zero vertical velocity.

The zero velocity contour encloses the vanishing cell as a vertical “tongue” at about $y = 4.0$ Mm in the upper left panel ($t = -2.0$ min). A downflow, originating from an intergranular lane, is deflected sideways into the deeper parts of the collapsing cell at about $z = -1.0$ Mm. As the cell collapses, the pressure decreases locally between $z = 0.0$ Mm and $z = 0.5$ Mm. This generates horizontal pressure gradients that accelerate the flow horizontally towards the site of collapse, and a horizontal mass flux convergence is initiated ($t = 0.0$ min). The pressure builds up due to the horizontal mass flux convergence ($t = 2.08$ min). The pressure support of the overlying atmosphere is reduced by the negative pressure perturbation between $z = 0.0$ Mm and $z = 0.5$ Mm, such that a downflow is initiated. This is reversed to upflow when the pressure increases ($t = 2.08$ min).

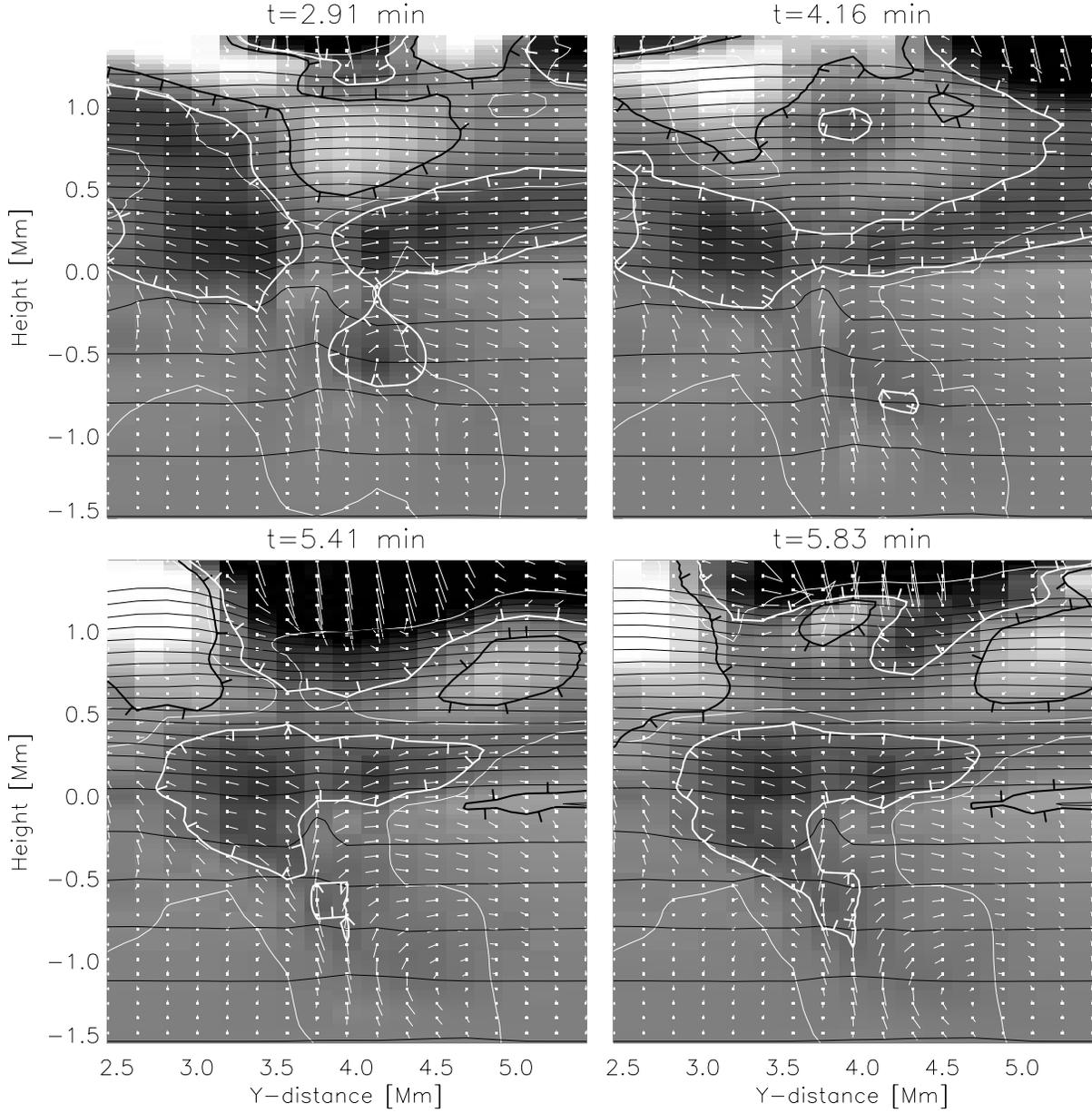


Figure 7: *Photospheric pressure perturbation and wave excitation, event (B)*. For the coding, see the previous Fig. 6. At $t = 2.91$ min, the pressure immediately above the downflow is still increasing, and a stagnation point has developed at about $y = 3.7$ Mm, $z = 0.2$ Mm. The upward velocity at $z = 0.7$ Mm is now maximum, with a corresponding pressure maximum, as in a propagating acoustic wave. At $t = 4.16$ min, this pressure maximum has induced horizontal outflow and corresponding mass flux divergence, with accompanying pressure decrease, as seen as the “flattened donut” centered at $y = 3.8$ Mm, $z = 0.8$ Mm. A horizontal wave component has now been induced.

At $t = 5.41$ min, gas flows towards this pressure minimum from the sides and from above. The downward moving gas falls into the relatively slow moving gas at around $z = 0.8$ Mm, resulting in non-linear steepening of an upward propagating wavefront (develops to a shock) at $z = 1.2$ Mm and $t = 5.83$ min. Here, we also see the horizontally propagating pressure perturbation, accompanied by raised density contours. This is a pressure modified gravity wave which is driven by buoyancy and horizontal pressure gradients.

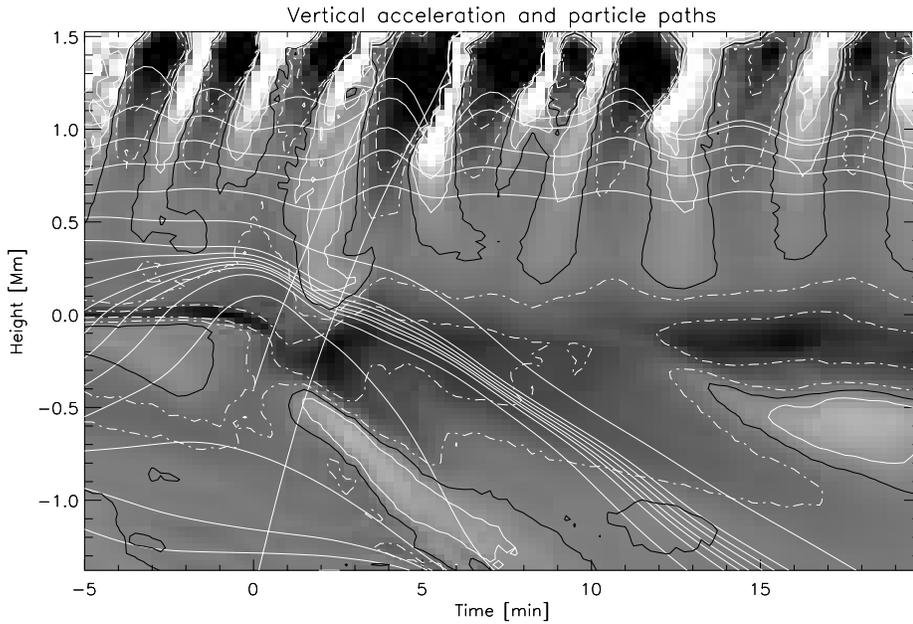


Figure 8: *Vertical acceleration, event (B)*. Vertical acceleration as a function of time along the vertical column centered on event (B). Upward acceleration corresponds to lighter shades of gray, and downward acceleration to darker shades. Zero acceleration is marked by black contours. White dashed contours mark -20 and -60 km s^{-2} (downward) accelerations, and white solid contours $+20$ and $+60 \text{ km s}^{-2}$ (upward) accelerations. The white, mostly horizontal lines, are particle paths obtained by time integration of the vertical velocity along the column. These have curvatures corresponding to the acceleration, as long as the horizontal velocity is negligible. The two white lines that cover the full height range indicate the sound velocity. The flow reversal in the cooling layer occurs at $t = 0.0$ min.

The downward moving structure in the convection zone is a signature of a plume, while the skewed contours in the atmosphere are wave fronts of gravity modified acoustic waves. Supersonic wave propagation is seen in the chromosphere (faster than the sound velocity line) at about $t = 5$ min and $z > 1.2$ Mm, due to a vertically propagating shock.

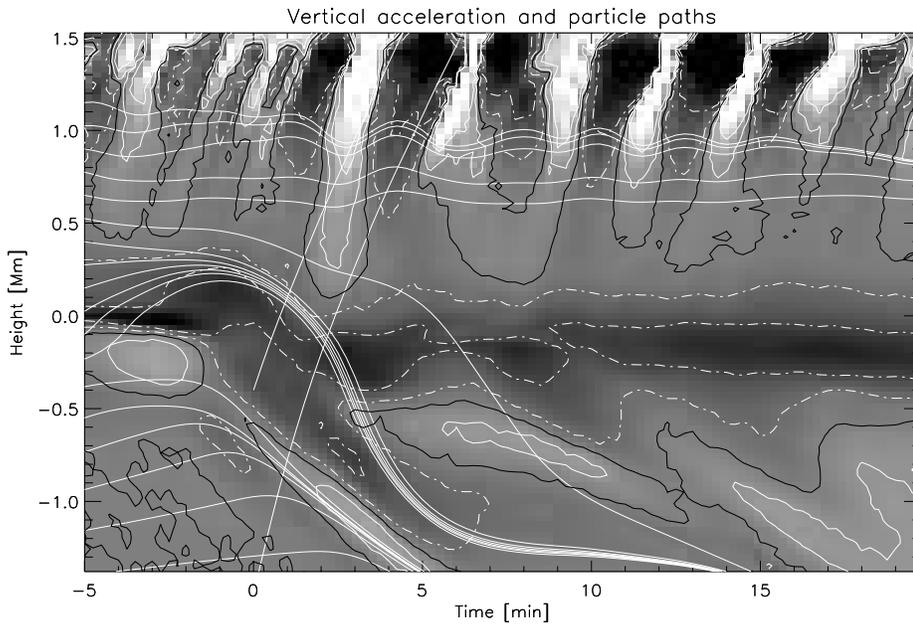


Figure 9: *Vertical acceleration, event (A)*. Vertical acceleration as function of time along the vertical column centered on event (A). The contours, curves and image have the same meaning as in Fig. 8. In this case, and in contrast to event (B), the atmospheric oscillation prior to the excitation is less coherent, and has less influence on the amplitude of the transient. The atmospheric response is in spite of this, similar to that of event (B). Note also the upward acceleration around $t = 2.0$ min that extends down to 100 km above the surface and into the region of wave excitation, as for event (B).

contribution from interference, qualitatively the same as for event (A).

For event (B) in chromospheric layers between $z = 1.0$ Mm and $z = 1.5$ Mm, the first downward motion associated with the transient reverses abruptly back to upward motion. This is a vertically propagating shock wave, and its supersonic propagation speed can be seen by a comparison with the sound velocity curves. The particle paths between subsequent shocks are similar to parabolas due to free fall under the pull of gravity. Supersonic propagation and shock formation occurs also for event (A), but in higher chromospheric layers at $t = 6.5$ min close to $z = 1.5$ Mm.

3.2.1 Convection dynamics prior to the granule collapse

Collapsing granules are predominantly located above subsurface downflows as discussed in Sect. 3.1.2. The convective upflow beneath the surface is in spite of this, accelerated upwards before the granule vanishes. The flow reversal in the granule is initiated in the same subsurface layers (not in the cooling layer) by a density increase, and gravity becomes stronger than the vertical pressure gradient.

Figs. 9 and 8 for events (A) and (B), show the initial upward acceleration below the surface between $z = -0.1$ Mm to about $z = -0.4$ Mm, from $t = -5.0$ min to about $t = -1.0$ min. After $t = -1.0$ min, upward acceleration reverses to downward, and at $t = 0.0$ min the upflow in the cooling layer reverses to downflow. The acceleration is downwards in deeper convective layers than $z = -0.5$ Mm before $t = 0.0$ min. This is also seen by the curvature of the particle paths. The flow is qualitatively similar also in subsurface layers, and I discuss event (B) in detail in the following.

Figure 10 shows the evolution of relative perturbations in pressure, vertical pressure gradient, and gravity along the column in the subsurface region. These perturbations are given by $P/\langle P \rangle - 1$, $(\partial P/\partial z)/\langle \partial P/\partial z \rangle - 1$ and $(g\rho)/\langle \partial P/\partial z \rangle - 1$ respectively, where the brackets denote time averaging. The gravity is normalized with the average pressure gradient, such that the figures can be compared directly in the sense that higher value of the pressure gradient perturbation than the gravity perturbation gives a net upward force. The black contour shows zero perturbations and the white contours, $\pm 5\%$ and $\pm 20\%$ perturbations.

The particle paths for the vertical motion (white lines) are diverging in subsurface layers before the collapse, indicating vertical mass flux divergence. Since only small fluctuations in density occur in the convection zone ($\nabla \cdot (\rho \vec{u}) \sim 0$), the accompanying vertical mass flux divergence is largely compensated by

horizontal mass flux convergence. This balancing is seen in Fig. 11, where I have displayed the relative time derivative of density, $(1/\rho)\partial\rho/\partial t$ (lower panel), and its contributions from horizontal (upper panel) and vertical (middle panel) mass flux convergence, $(-1/\rho)\nabla_h \cdot (\rho \vec{u}_h)$ and $(-1/\rho)\partial(\rho u_z)/\partial z$ respectively. The horizontal mass flux convergence is due to overturning fluid from neighboring upflow. This is consistent with the previous result that collapsing granules are located above subsurface downdrafts with horizontally convergent flow. The density is increasing in time around $z = -0.5$ Mm, as seen by the positive time derivative (lower panel). This is due to excess horizontal mass flux convergence.

A positive pressure perturbation develops around $z = -0.5$ Mm (left panel in Fig. 10), that is larger than the pressure perturbation above and below this level. This is mainly due to the increasing density (adiabatic changes) at this level. This localized pressure perturbation leads to an enhanced pressure gradient towards the cooling layer as seen in the middle panel in Fig. 10. The gravity also increases due to increased density (right panel in Fig. 10), but this is not large enough to compensate for the increased pressure gradient at this stage. This is seen by comparing the contours in the middle and right panels.

As a result, the upward acceleration is enhanced in the upflow immediately beneath the cooling layer, and the upflow velocity in the cooling layer increases⁷. This was also noted in the velocity curves for the low viscosity events in Fig. 2. Note that the collapse is initiated in subsurface layers by horizontally converging flow that leads to larger upflow velocity in the granule. If, in contrast, the collapse was initiated in the surface by radiative over-cooling, we would expect monotonically decreasing upflow velocity as in the central parts of an “exploding” granule.

The granule collapses around $t = 0.0$ min after downward acceleration has been initiated below the surface at $t = -1.0$ min (see $z = -0.1$ Mm to $z = -0.4$ Mm in Fig. 8). This happens when gravity becomes larger than the pressure gradient (Fig. 10). This is also seen as negative curvature of the particle paths. The increased density is again due to excess horizontal mass flux convergence, as compared to the vertical divergence (Fig. 11).

⁷This is opposite to “buoyancy braking” of the upflow in large, expanding granules, in which the positive pressure perturbation in the surface lowers the vertical pressure gradient, and brakes the upflow. In the current case, “buoyancy acceleration” occurs above a positive pressure perturbation in deeper layers.

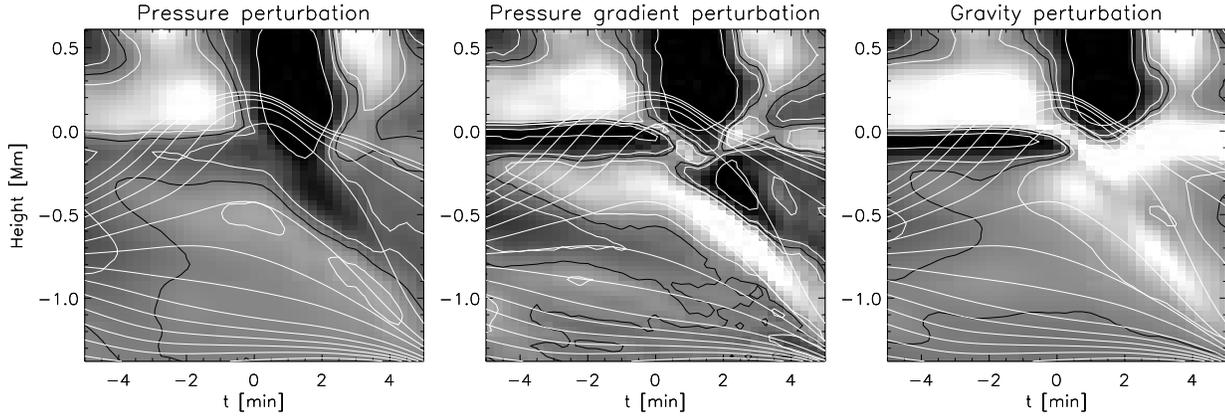


Figure 10: *Relative perturbations in pressure, vertical pressure gradient and gravity prior to collapse, event (B).* The contours show zero (black), $\pm 5\%$ and $\pm 20\%$ (white) relative perturbations. The particle paths for vertical motion are also shown.

The relative pressure perturbation (left panel) increases in time in subsurface layers partially due to mass flux convergence from overturning fluid from neighboring upflow. This raises the pressure gradient (middle panel) towards the surface above $z = -0.7$ Mm, and serves to increase the upflow velocity in the cooling layer. Diverging particle paths correspond to vertical mass flux divergence, and this is slightly over-compensated by the horizontal mass flux convergence. Therefore, the gravity increases (right panel), but that is not sufficient to compensate for the raised pressure gradient close to the surface, until about $t = -1.0$ min. The pressure gradient and gravity images with contours are directly comparable since the gravity has been normalized the same way as the pressure gradient.

3.2.2 Wave excitation in the overshoot zone

The signature of wave excitation was seen for events (A) and (B) in Figs. 9 and 8 (at $t \sim 2.0$ min), as a propagating region of upward acceleration from the lower photosphere and upwards. The slanted contour which mark zero acceleration in the figures delineates this region and show upward acceleration down to $z = 0.1$ Mm. Below that level, the convective flow adds a negative acceleration component, but this is perturbed by a smaller upward acceleration during wave excitation. The contours which mark negative acceleration in the figures indicate that this perturbation extends down to the cooling layer.

If the wave source was located in the convection zone, we would have seen this perturbation also in deeper layers. I therefore conclude that the wave source is not located in the convection zone, but in the convective overshoot zone (from the cooling layer at $z = 0.0$ Mm and to about 0.5 Mm above).

The structures marked by the slanted contours in the convection zone after $t = 0.0$ min (Figs. 9 and 8) are a signature of subsonic downward motion of the cold plume, and not downward propagating acoustic waves from the source region. The motion is subsonic as seen by a comparison with the sound velocity curves.

In the following, I concentrate on the dynamics in the overshoot zone that is responsible for excitation of the vertical wave components in the atmosphere. I use event (B) for this study.

The vertical acceleration that generates vertical wave components is driven by fluctuations in the pres-

sure gradient and in density (gravity). To explain the vertical wave excitation, I therefore consider the sources of these pressure and density fluctuations. The changes in horizontal flow are also discussed, since these give a strong influence on the pressure and density fluctuations.

First, I describe the density fluctuations by considering horizontal and vertical mass flux divergence. Having found the sources of the density fluctuations, I can discuss the adiabatic pressure fluctuations that are controlled by these density changes. Then, I explain the non-adiabatic pressure fluctuations in terms of heat advection and radiative heating.

I decompose the pressure fluctuations in adiabatic and non-adiabatic components (cf. Sect. 2.2), to find their relative importance. The non-adiabatic component is not as large as the adiabatic component, such that the total pressure is mainly controlled by adiabatic changes. However, the non-adiabatic contribution is significant and serves to damp the pressure fluctuations.

After an explanation of the pressure and density fluctuations, I am in a position to discuss the differences between vertical pressure gradient and gravity that drives the wave excitation. The vertical pressure gradient is also decomposed in adiabatic and non-adiabatic contributions, in order to determine their effect on the wave excitation.

Density fluctuations. The density is decreasing in time in the overshoot zone, in the interval $t = [-1.5, 1.5]$ min (before and after the cell collapse),

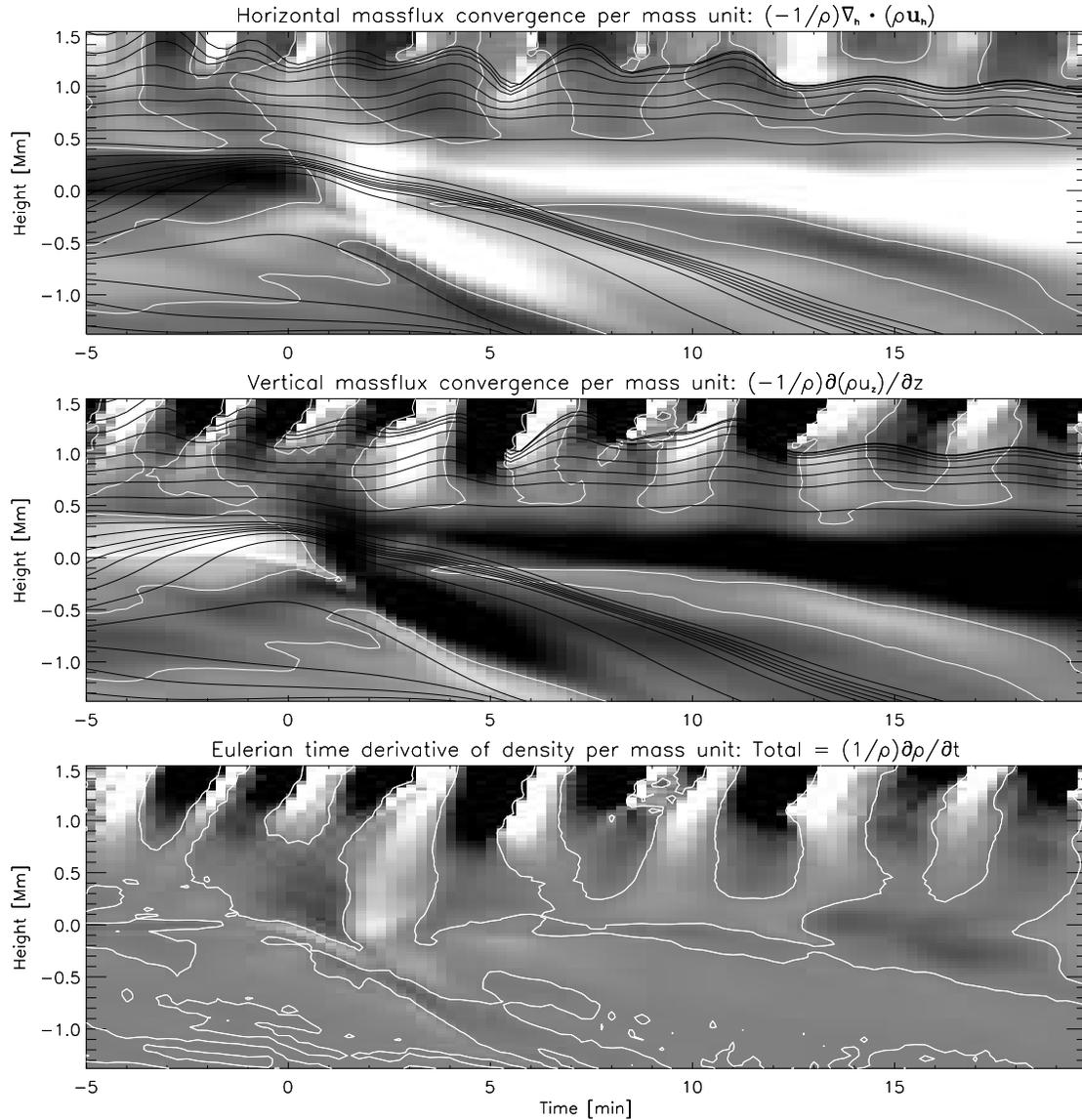


Figure 11: *Mass conservation, event (B)*. The images show the time evolution along the vertical column centered on the event, of horizontal (upper panel) and vertical (middle panel) mass flux convergence, normalized by density. The lower panel shows the sum of these contributions, which is the relative rate of change of density. The time axis shows elapsed time in minutes after downward motion starts in the granule at $z=0.0$ Mm. White contours mark zero mass flux divergence, and zero rate of change of density. The black lines are particle paths corresponding to vertical velocity along the column.

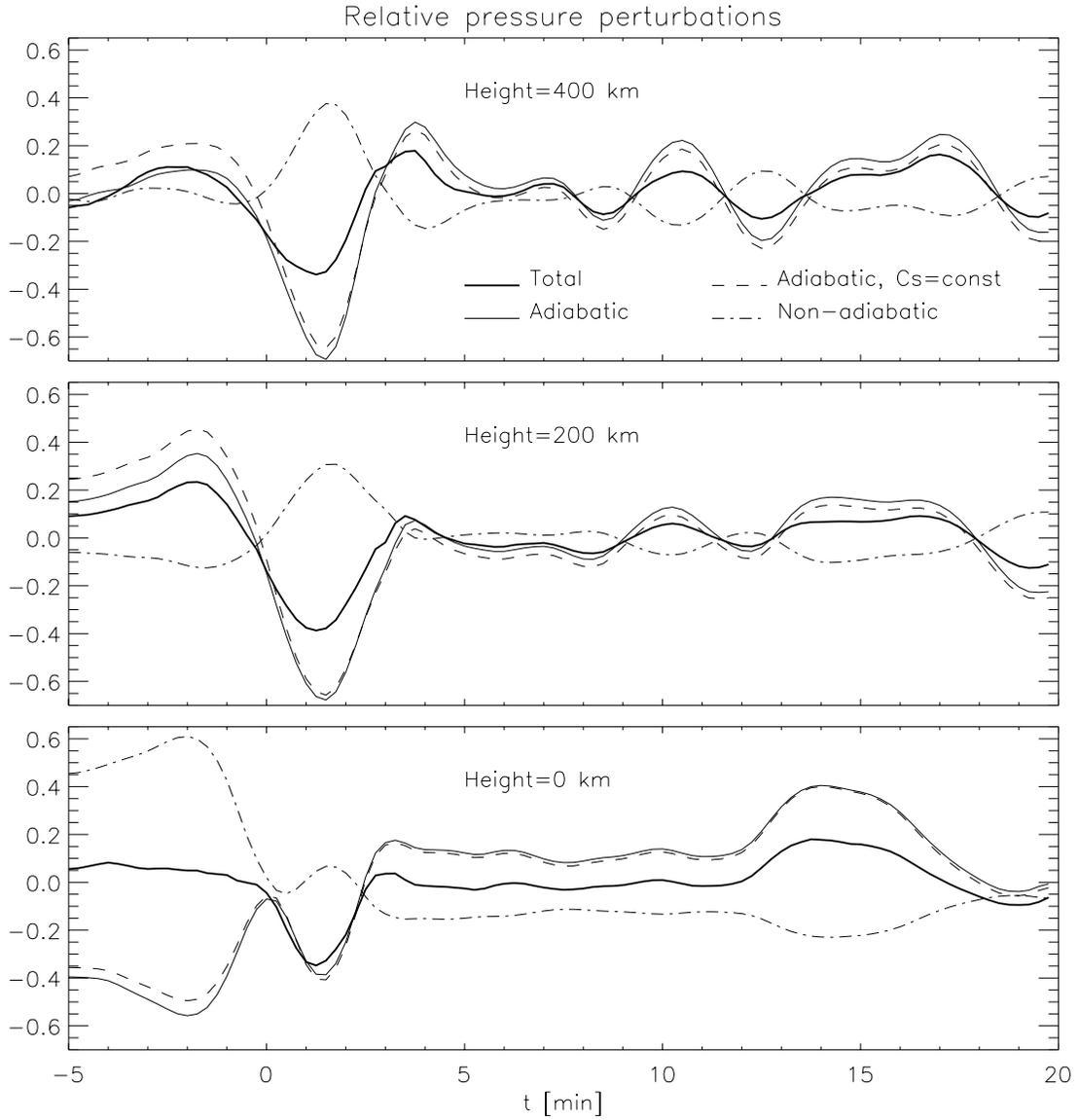


Figure 12: *Pressure perturbations in the overshoot zone, event (B)*. These perturbations are given by the difference between the actual pressures and the corresponding time averages, and divided by the total time averaged pressure. The total pressure perturbation is the sum of adiabatic and non-adiabatic contributions.

The actual adiabatic pressure perturbation does not deviate very much from the adiabatic pressure fluctuation resulting from a constant sound velocity (essentially temperature). The total pressure perturbation is mainly controlled by the adiabatic fluctuation, but the amplitude is lower than the adiabatic fluctuation, due to non-adiabatic damping. The lower panel shows the perturbations in the cooling layer. All perturbations are measured directly above the collapsing granule (actually within it at 0.0 km).

as seen by the negative time derivative in Fig. 11. In the upper and middle panels, we see that this is initially due to excess horizontal mass flux divergence (divergent horizontal overshoot flow), compared to vertical mass flux convergence (granular upflow). Around $t = 1.0$ min, the vertical mass flux becomes divergent, due to the newly formed downdraft. The horizontal flow becomes convergent (due to the pressure deficiency discussed below), but the resulting horizontal mass flux convergence can at this stage not compensate for the vertical divergence, and the density continues to decrease.

Eventually, at about $t = 1.5$ min, the horizontal mass flux convergence over-compensates the vertical divergence, and the density increases again.

Pressure fluctuations and driving of horizontal flow. Fig. 12 shows the relative pressure fluctuations at two different levels in the overshoot zone, at $z = 200$ km and $z = 400$ km, and in the cooling layer at $z = 0.0$ km. The total pressure is split into an adiabatic part and a non-adiabatic part, and the fluctuations in these quantities relative to the total pressure are shown in the figure. The two pressure contributions are calculated by a time integration of Eq. (3), and they are given by:

$$P_{\text{adiab.}}(t) = \int \left\{ c_s^2 \frac{\partial \rho}{\partial t} \right\} dt + \text{const.}$$

$$P_{\text{non-adiab.}}(t) = \int \left\{ (\Gamma_3 - 1) \rho T \frac{\partial s}{\partial t} \right\} dt + \text{const.}$$

The fluctuations are calculated by subtracting the time averaged values (thereby eliminating the constants of integration), and dividing by the time averaged total pressure, e.g., $(P_{\text{adiab.}} - \langle P_{\text{adiab.}} \rangle) / \langle P_{\text{Total}} \rangle$. Adiabatic fluctuations are shown by the thin, solid line and non-adiabatic fluctuations with the thin dash-dotted line. The total pressure fluctuation is the sum of the two contributions, and is shown by the solid, thick line.

At $z = 200$ km and $z = 400$ km, the total pressure fluctuation is mainly dominated by, and therefore in phase with, the adiabatic fluctuation. The non-adiabatic contribution is in anti-phase with the adiabatic fluctuation, and that gives an amplitude damping of the total pressure fluctuation, relative to the adiabatic.

The squared sound velocity is essentially proportional to the temperature, and can modulate the influence from the density changes. The dashed curve shows the adiabatic fluctuation that would result for constant sound velocity. I have here used the time averaged sound velocity at the same levels. This curve is not radically different from the actual adiabatic fluctuation, and we can therefore conclude that the tem-

perature modulation is not very important. This can be understood, since the photospheric gas is kept close to the radiation temperature (of the photons from the cooling layer below) by radiative heating or cooling.

Hence, we can conclude that the pressure changes in the overshoot zone are mainly controlled by density changes, and that there is an amplitude damping from non-adiabatic effects. The pressure change is initially driven by the reduction in the vertical mass flux convergence (upflow at this stage), and the pressure drops locally mainly due to reduced density. The resulting pressure deficiency (lower pressure than the surroundings) brakes the horizontal outflow by the horizontal pressure gradients that are set up, and at $t = 1.0$ min, the divergent horizontal flow has reversed to convergent flow. The delayed response of the horizontal flow to the decreasing pressure is due to fluid inertia.

Later, at $t = 1.5$ min, the resulting horizontal mass flux convergence is able to balance the vertical mass flux divergence above the newly formed downdraft. After that, the density and therefore the pressure is increasing, due to continued excessive horizontal mass flux convergence. A pressure maximum relative to the surroundings at the same height is now building up.

A temporary stagnation point, with upflow above and downflow below, is formed around the pressure maximum at about $t = 3.0$ min and $z = 200$ km (see Fig. 7 for $t = 2.91$ min). The density and pressure reach a maximum at $t = 3.5$ min, but this is not sufficient to set up horizontal pressure gradients that are able to halt the horizontal inflow. The flow pattern is upheld, with horizontal mass flux convergence and vertical mass flux divergence, and relatively small pressure perturbations in the overshoot zone. This is the normal situation in the overshoot zone above a downdraft.

Non-adiabatic contributions to pressure changes. The non-adiabatic contribution to the pressure derivative $\partial P / \partial t$ is split in a heat advection term and a radiative/viscous heating term. The evolution of these terms, along the column centered on the event, is shown in Fig. 13. The quantities that are displayed are calculated as follows.

The non-adiabatic contribution is (from Eq. 3):

$$(\Gamma_3 - 1) \rho T \frac{\partial s}{\partial t},$$

and is shown in the middle left panel. The rate of change of heat, $T \partial s / \partial t$ is (from Eq. 4):

$$T \frac{\partial s}{\partial t} = q_{\text{rad.}} + q_{\text{visc.}} - \vec{u} \cdot T \nabla s,$$

where $q_{\text{rad.}}$ and $q_{\text{visc.}}$ are radiative and viscous specific heating rates respectively. We can therefore split the

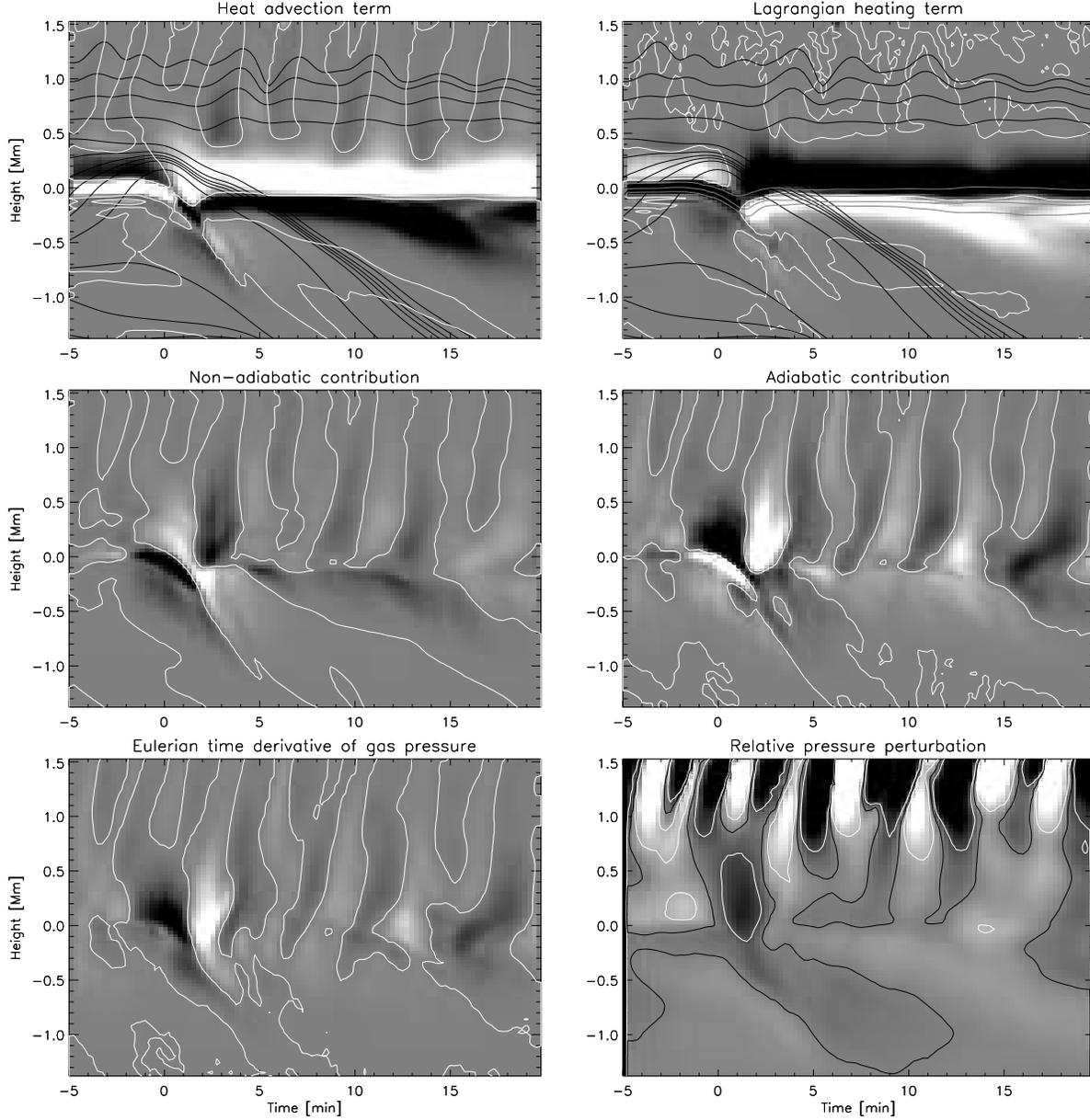


Figure 13: *Contributions to non-adiabatic pressure changes, event (B).* All panels show the time evolution along the column centered on the event. The sum of the heat advection term (upper left panel) and the contribution from radiative energy exchange and viscous heating (upper right panel), is the rate of change of pressure due to non-adiabatic sources (middle left panel). The adiabatic contribution is in the middle right panel. The sum of the contributions in the middle panels is the total rate of change of pressure $\partial P/\partial t$ (lower left panel). The white contours in these figures mark the zero-level. The black contours in the upper two panels are particle paths obtained from time integration of the vertical velocity. These are shown to indicate the direction of heat advection. The grey, mostly horizontal contours in the upper right panel are the levels of vertical optical depths of $\tau = 1, 10, 50$. The contours show the vertical displacements of the cooling layer. The lower right panel is the relative pressure perturbation (relative to the time average at each height). The black contour show zero perturbation, and the white contours show $\pm 20\%$ perturbations. All of the quantities, except for the relative pressure perturbation, are arbitrarily scaled with e^{6z} to show variations at a large height range.

The pressure is mainly controlled by adiabatic changes in the atmosphere. The non-adiabatic changes are in anti-phase with the adiabatic changes, and serve to damp the adiabatic effect on the total pressure changes.

non-adiabatic contribution into a heat advection term and a Lagrangian heating term. The heat advection term is:

$$-\rho(\Gamma_3 - 1)\vec{u} \cdot T\nabla s.$$

This is shown in the upper left panel⁸. The Lagrangian heating term is :

$$\rho(\Gamma_3 - 1)(q_{\text{rad.}} + q_{\text{visc.}}).$$

This is shown in the upper right panel. A positive contribution from these heating terms can balance expansion work (a negative adiabatic contribution), and/or increase the pressure. The sum of the two terms in the upper panels is equal to the non-adiabatic contribution in the middle left panel.

The middle right panel shows the adiabatic contribution:

$$c_s^2 \frac{\partial \rho}{\partial t}.$$

The lower left panel shows $\partial P/\partial t$, which is the sum of the middle panels, and the lower right panel shows the relative pressure perturbation, that is essentially the time integral of $\partial P/\partial t$. All of these terms have been scaled by an arbitrary factor e^{6z} to show variations at a large height range in the images.

Non-adiabatic pressure changes in the overshoot zone. The overshoot flow at $z = 200$ km is radiatively heated before the collapse (see Fig. 13), and this is almost balanced by negative heat advection due to the upward velocity (specific entropy increasing upwards). After $t = 0.0$ min, the downflow has started, and the heat advection becomes positive. This is under-compensated by radiative cooling, such that the non-adiabatic pressure increases in time. This is the reason for the increasing non-adiabatic contribution in Fig. 12. After $t = 2.0$ min, radiative cooling is larger than the positive heat advection, and the non-adiabatic contribution to the pressure decreases. This is the reason for the decreasing non-adiabatic contribution in Fig. 12.

In higher layers at $z = 400$ km, the non-adiabatic contribution is mainly due to the vertical heat advection, with increasing (decreasing) non-adiabatic pressure for the downward (upward) moving phases. Radiative cooling is in anti-phase, and serves to damp the advection effect.

⁸The specific entropy s decreases with height in the cooling layer (convective instability) and increases with height in the overlying atmosphere (convectively stable). The heat advection is therefore negative for upward motion in the stable atmosphere.

Non-adiabatic pressure changes in the cooling layer. There is a large positive contribution from upward heat advection in the cooling layer at $z = 0.0$ km (see Fig. 13). This is almost balanced by radiative cooling, but initially, there is some net heating and increasing non-adiabatic pressure, due to the increasing upflow velocity just before the collapse. The increasing non-adiabatic pressure is almost cancelled by a negative contribution from decreasing density (see also the bottom panel in Fig. 12), and the total pressure remains almost constant. The cooling layer therefore expands isobarically due to net heat input.

At about $t = -2.0$ min, the reduced upward heat advection is not sufficient to balance the radiative cooling, and the non-adiabatic pressure decreases. This is again compensated by the adiabatic contribution that is positive, due to increasing density. This is isobaric contraction due to net heat loss.

Hence, due to this cancellation effect, the total pressure perturbation in the cooling layer is not very large nor very rapid ($\sim 10\%$ in 5 min), prior to the flow reversal in the granule. This is shown in the bottom panel in Fig. 12.

A negative pressure perturbation ($\sim -30\%$) occurs after the flow reversal, but this is mainly due to adiabatic changes. The pressure is decreasing mainly due to adiabatic expansion in the downdraft, and then it increases by a following adiabatic compression. The pressure changes that contribute to wave excitation in the cooling layer (which is the base of the overshoot zone) are therefore not primarily caused by thermodynamic heat sources, but by adiabatic changes due to imbalance between vertical and horizontal mass flux divergence.

The upper right panel in Fig. 13 shows that the cooling layer moves down in the initial phases, as seen by the shift of the optical depth curves, $\tau = 1, 10, 50$. Note that as this surface starts to move down, gas is flowing upwards through it. After the downdraft has formed, radiation from surrounding granules heats the downflow, and the optical depth curves are shifted slightly upwards again.

Force balance and vertical acceleration in the overshoot zone. In the initial stages of the wave excitation, the acceleration is driven by changes in the adiabatic pressure gradient and in gravity. First, I therefore discuss the response to density changes in the adiabatic pressure gradient and compare that to the response in gravity. In later stages of the wave excitation, the non-adiabatic contribution to the pressure gradient becomes important. This effect must then be included in the discussion to explain the perturbations in the total vertical forcing and acceleration.

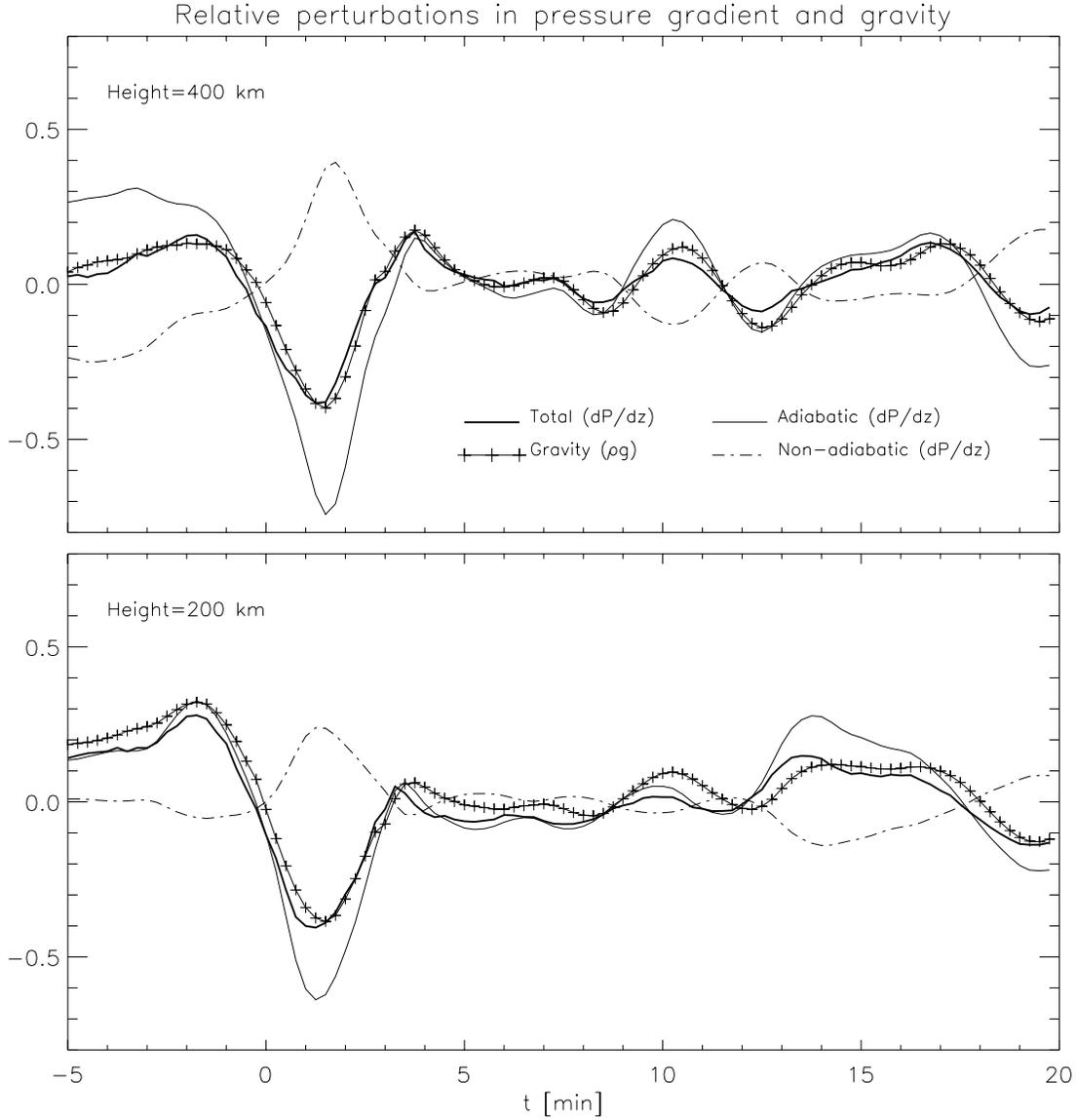


Figure 14: *Vertical force balance in the overshoot zone, event (B).* The perturbations in vertical pressure gradient and gravity have been normalized (see the text) by the time averaged pressure gradient at each height, such that when the gravity curve is larger than the pressure gradient, downward acceleration results. The sum of the adiabatic and non-adiabatic perturbations is equal to the total pressure gradient perturbation.

Downward acceleration starts at 400 km just before the granule collapses at $t = 0.0$ min, mainly due to adiabatic reduction of the pressure gradient (caused by reduced density), which dominates over the corresponding reduction in gravity. The increasing non-adiabatic contribution (by positive downward heat advection) eventually raises the pressure gradient above the gravity, and upward acceleration starts. The adiabatic pressure gradient now increases (due to horizontal mass flux convergence) at a larger rate than the now decreasing non-adiabatic contribution (by negative upward heat advection), such that the upward acceleration continues. At 200 km, the pressure gradient is initially lower than the gravity due to balancing of upwards decreasing vertical momentum flux in the convective flow.

Fig. 14 shows the relative fluctuation in the pressure gradient, $(dP/dz - \langle dP/dz \rangle) / \langle dP/dz \rangle$, and the normalized fluctuation in gravity, $(g\rho - \langle dP/dz \rangle) / \langle dP/dz \rangle$. The normalization factor is the time averaged pressure gradient at the given height. Using the same normalization makes it possible to determine the direction of the acceleration by a direct comparison of the curves. The contributions from adiabatic and non-adiabatic pressure gradients are also shown in the figure. These contributions are given by a time integration of Eq. (6), and normalized such that their sum is equal to the relative fluctuation in the total pressure gradient.

The adiabatic pressure gradient falls faster than the gravity in the phase of downward acceleration (where the gravity is larger than the total pressure gradient). Later, in the phase of positive acceleration, the adiabatic pressure gradient rises faster than the gravity. By using the adiabatic part of Eq. (6), we can write this effect as:

$$g \left| \frac{\partial \rho}{\partial t} \right| < \left| \frac{\partial}{\partial z} \left\{ c_s^2 \frac{\partial \rho}{\partial t} \right\} \right| = \left| \frac{\partial}{\partial z} \left\{ c_s^2 \nabla \cdot (\rho \vec{u}) \right\} \right|. \quad (18)$$

From the density perturbations in Fig. 11, we see that for a change $\delta\rho$ in time, $\delta\rho/\rho \sim \text{const.}$ in the height direction in the overshoot zone. We can therefore write $\delta\rho \sim e^{-z/H}$, since the density varies approximately exponentially with height. H is the density scale height. We can then write the last equation as (setting the sound velocity to a constant, typical value):

$$g |\delta\rho| < \left| \frac{c_0^2}{H} \delta\rho \right|. \quad (19)$$

This corresponds to a Froude number⁹ $F_r = c_0^2/H/g > 1$. In the current case for the overshoot zone, $F_r \sim 2$, using typical values in this layer. Hence, the rate of change of the adiabatic pressure gradient is only marginally larger than the rate of change of gravity.

The change in the total pressure gradient is also influenced by the non-adiabatic contribution in Eq. (6):

$$\frac{\partial}{\partial z} \left\{ (\Gamma_3 - 1) \rho T \frac{\partial s}{\partial t} \right\}.$$

For the purpose of demonstration, I assume that a change in time of heat, $\delta q = T\delta s$, is constant with height¹⁰. I also set Γ_3 constant with height, which is a

⁹The Froude number is also the typical ratio (inertial term)/(gravity term) in the equation of motion.

¹⁰This approximation is valid when the advection part includes a linear variation of s with height (which is a good approximation), and constant vertical velocity. We must also assume constant specific radiative heating with height.

good approximation in the overshoot zone. The non-adiabatic contribution now becomes, with $\rho \sim e^{-z/H}$:

$$(\Gamma_3 - 1) \frac{\delta Q}{H}, \quad (20)$$

for a change of heat per volume unit δQ [erg cm⁻³]:

$$\delta Q = \rho \delta q = \rho (q_{\text{rad.}} + q_{\text{visc.}} - \vec{u} \cdot T \nabla s). \quad (21)$$

This example shows that the pressure gradient increases by radiative and viscous heating, and/or positive heat advection by downward motion in the atmosphere.

Gathering all contributions, I can now write, using Eq. (1), an approximate equation for the perturbation in the vertical force:

$$\delta F_z \sim \left\{ \frac{c_0^2}{H} - g \right\} \delta\rho + (\Gamma_3 - 1) \frac{\delta Q}{H}. \quad (22)$$

Then, consider Fig. 14. Downward acceleration is initiated at 400 km ($t = -1.5$ min) by a reduction in the adiabatic pressure gradient. As the density reduces due to reduced vertical mass flux convergence (and sustained horizontal divergence), the pressure gradient falls adiabatically, and at a higher rate of change than the gravity. The non-adiabatic modification contributes little to the total pressure gradient in the initial phase, and downward acceleration results.

The downward acceleration initiates downward velocity and an increase in the non-adiabatic pressure gradient, due to positive heat advection. This is because the specific entropy increases in the upward direction. This damps the influence from the decreasing adiabatic pressure gradient. At $t = 1.5$ min, the positive contribution from the non-adiabatic pressure gradient serves to raise the total pressure gradient above the gravity, and upward acceleration starts.

Immediately afterwards, the downflow reverses to upflow, and the contribution from the non-adiabatic pressure gradient decreases, due to negative heat advection. In spite of this, the total pressure gradient continues to increase because the rise in the adiabatic pressure gradient overcomes the fall in the non-adiabatic pressure gradient. The adiabatic pressure gradient is rising due to excess horizontal mass flux convergence. The rate of change in the adiabatic pressure gradient is again larger than the rate of change in gravity.

Maximum upward velocity is reached at 400 km and above, when downward acceleration eventually sets in again (Fig. 8). The atmosphere continues to oscillate vertically, with a period close to the cut-off period of 3 min.

The fluctuations in pressure gradient and gravity for 200 km are qualitatively the same as for 400 km in the stages of wave excitation ($t = -1.5$ min to

$t = 3.0$ min), as seen in Fig. 14. The fluctuations in net vertical force are therefore also qualitatively the same, and caused by the same processes.

Before $t = -1.5$ min, we see that the 400 km curves show approximate hydrostatic equilibrium, while the 200 km curves do not. The gravity is somewhat larger than the pressure gradient for 200 km, with resulting net downward force. This is not a contribution to the wave excitation, but the contribution from momentum balance in the convective flow prior to the collapse. This contribution is stronger at heights closer to the cooling layer. In steady state (slowly evolving convective flow prior to cell collapse), and assuming that horizontal velocities are small, we can write this contribution as:

$$\rho \frac{\partial}{\partial z} \left\{ \frac{1}{2} u_z^2 \right\} = - \frac{\partial P}{\partial z} - g\rho < 0. \quad (23)$$

The inequality sign stems from the fact that the magnitude of the vertical velocity is decreasing with height in both upflows and downdrafts in the normal, slowly evolving convective flow.

4 Summary

Simulations of solar convection and the overlying atmosphere are examined to find sources of acoustic waves, and how these waves interact with the atmosphere. It is found that atmospheric wave transients are excited in the convective overshoot zone by vanishing/collapsing granules. These granules are relatively small and stem from the splitting of larger granules.

Splitting of granules produces granule fragments of different sizes. The splitting process is well understood by the effect of buoyancy braking and subsequent radiative over-cooling, and growth of regions with dense cool material within the granules (e.g., Nordlund 1985). The larger fragments expand, while the smaller fragments eventually vanish or collapse, in the sense that upflow reverses to downflow and leaves behind a downdraft. This work shows that collapsing granules are preferentially located above subsurface downdrafts extending to depths of up to 1000 km below the cooling layer. This suggests that collapsing granules and the associated wave excitation are separated by a typical length scale on the solar surface, corresponding to mesogranular scales of the convective flow beneath the surface.

The granule collapse is initiated in the subsurface layers by horizontally converging flow near the downdrafts, on the order of 10 min prior to the flow reversal in the cooling layer. The lower parts of the convective upflow associated with the granule are subject to a net downward force, with subsequent downward motion. The upper parts are subject to an enhanced pressure

gradient towards the cooling layer, and that leads to increasing upflow velocity in the cooling layer prior to the collapse. This is quite different from a scenario in which over-cooling in the cooling layer serves to decrease the upflow velocity monotonically in time, as in the central parts of a granule that splits into several fragments (“exploding” granule).

Upflow reverses to downflow in the cooling layer (the granule vanishes) after downward acceleration has been initiated below the surface. Downward acceleration is initiated when gravity (density) eventually becomes larger than the pressure gradient. The increased density is due to excess horizontal mass flux convergence from neighboring granules.

A negative pressure perturbation is generated in the convective overshoot zone in response to the granular flow reversal. This zone is located in the convectively stable atmosphere, and extends from the cooling layer and to 500 km above. The maximum amplitude of the pressure perturbation is $\sim -30\%$ relative to the unperturbed pressure, and the fall- and rise-times are approximately 2 minutes, which is slightly shorter than the cutoff period for vertical acoustic wave propagation. This “implosion” in the overshoot zone generates a wave transient with dominating vertical wave components.

The pressure perturbation is mainly driven by density fluctuations (adiabatic changes), but non-adiabatic heating and cooling (heat advection and radiative heating) damps this effect. The decreasing pressure is initially due to decreasing density caused by smaller vertical mass flux convergence in the granular upflow, than the horizontal mass flux divergence. The horizontal outflow reverses to inflow due to the localized pressure deficiency that sets up horizontal pressure gradients. The horizontal mass flux becomes convergent, and over-compensates for the vertical mass flux divergence immediately after the downflow has been initiated. This serves to increase the pressure to about the initial value.

It is interesting to note that the pressure perturbation in the cooling layer is not very large nor very rapid ($\sim 10\%$ in 5 min), prior to the flow reversal in the granule. This is due to an approximate cancelation between the adiabatic and non-adiabatic perturbations. The gas cools by radiation and contracts isobarically. A negative pressure perturbation ($\sim -30\%$), that is in phase with the perturbation in the overshoot zone, occurs after the flow reversal, but this is mainly due to adiabatic changes.

The vertical acceleration that drives the vertical wave component, is initially downwards because the vertical pressure gradient drops more rapidly than gravity. In this phase, the pressure gradient is mainly controlled by density changes. The following upward

acceleration is initiated by downward heat advection and radiative heating that damps the adiabatic reduction of the pressure gradient. The upward acceleration continues because the pressure gradient increases faster than the increasing gravity. The rising pressure gradient is in this phase mainly due to increasing density caused by excess horizontal mass flux convergence from the overshoot flow from neighboring granules. Upward velocity eventually results, and the atmospheric oscillation is initiated.

The oscillation in vertical velocity at 700 km above the surface, shows a delayed maximum in upward velocity after the cell collapse. The time delay spread is in the range 1 – 3 min. The peak velocity for the oscillation lies in the range 1 – 2 km s⁻¹ with an average peak velocity of 1.25 km s⁻¹. The envelope of the 3 min oscillation grows from $t = 0.0$ min (when the cell collapses) and reaches the peak at $t = 5.0$ min, after which it decays. In contrast, the envelope of the 5 min band decays immediately after $t = 0.0$ min. This indicates that energy is fed into the 3 min eigenoscillation of the atmosphere rather than the 5 min p-mode oscillation.

I find upward directed acoustic wave energy flux (Pu_z -flux) in the atmosphere above the vanishing granules. This is accompanied by a darkening in the granular (continuum) intensity as the small granule vanishes. This is supported by the observations of Rimmele et al. (1995), in which upward directed wave energy flux is accompanied by darkening of intergranular lanes. Furthermore, Espagnet et al. (1996) observed darkening in the granulation with an accompanying transient intensity oscillation that also supports the simulation results.

Some types of “internetwork bright grains” in the CaII H and K line cores (Rutten & Uitenbroek 1991 and Lites et al. 1993) are generated by vertically propagating shock trains in the chromosphere as shown by the 1D simulations of Carlsson & Stein (1997). This work shows that collapsing granules produce vertically propagating wave components that steepen into shocks in chromospheric layers above $z = 1.0$ Mm. Collapsing granules can therefore be a wave source for internetwork bright grains, where magnetic fields are weak.

In addition to excitation of vertically propagating, gravity modified acoustic waves, I have also detected excitation of pressure modified gravity waves. These waves are not detectable for more than 2-3 oscillation periods, since the amplitude is damped by horizontal expansion, and they interfere with the external atmospheric wave field.

Acknowledgements

Robert F. Stein and Åke Nordlund are thanked for letting me build onto their convection code. Robert F. Stein is thanked for suggestions on changes to the manuscript. Andreas Botnen is thanked for his conversion of the new radiation part in the code to parallel versions running on different parallel computers. Nick Hoekzema and Peter Brandt are thanked for illuminating discussions regarding the real sun versus simulations. Rob Rutten is thanked for help with improving the outline of this paper, and for providing observational data of the sun. Øyvind Andreassen at Forsvarets Forskningsinstitut (FFI) in Norway, and the group at Colorado Research Associates (CoRA) in Boulder, USA, are thanked for providing efficient volume visualization tools, and for introducing me to vortex dynamics, and for suggestions on manuscript improvements. Mark Rast at the HAO is thanked for his interest in my work, and for manuscript reading. Mats Carlsson is thanked for manuscript reading and for being my Ph.D. advisor. Teresa Lynne Palmer is thanked for manuscript reading.

This work has received support from The Research Council of Norway through a Ph.D. grant, and through a grant of computing time (Program for super-computing).

References

- Bracewell R. N. 1986, *The Fourier Transform and Its Applications*, McGraw-Hill, Inc.
- Carlsson M., Stein R. F. 1992, *ApJ*, 397, L59
- Carlsson M., Stein R. F. 1994, in M. Carlsson (ed.), *Chromospheric Dynamics*, Proc. Miniworkshop, Inst. Theor. Astrophys., Oslo, p. 47
- Carlsson M., Stein R. F. 1997, *ApJ*, 481, 500
- Cheng Q.-Q., Yi Z. 1996, *A&A*, 313, 971
- Crighton D. G. 1981, *J. Fluid Mech.*, 106, 261
- Espagnet O., Muller R., Roudier T., Mein P., Mein N., Malherbe J. M. 1996, *A&A*, 313, 297
- Goldreich P., Kumar P. 1990, *ApJ*, 363, 694
- Gustafsson B. 1973, *Uppsala Astr. Obs. Ann.*, 5, No. 6
- Gustafsson B., Bell R. A., Eriksson K., Nordlund Å.. 1975, *A&A*, 42, 407
- Howe M. S. 1975, *J. Fluid Mech.*, 71, 625
- Lamb H. 1908, *Proc. London Math. Soc.*, 7, 122
- Lighthill M. J. 1952, *Proc. R. Soc. Lond. A*, 211, 564
- Lites B. W., Rutten R. J., Kalkofen W. 1993, *ApJ*, 414, 345
- Musielak Z. E., Rosner R., Stein R. F., Ulmschneider P. 1994, *ApJ*, 423, 474
- Nordlund Å. 1985, in H. U. Schmidt (ed.), *Theoretical Problems in High Resolution Solar Physics*,

- MPA/LPARL Workshop, Max-Planck-Institut für Physik und Astrophysik MPA 212, München, p. 1
- Nordlund Å., Stein R. F. 1990, *Comp. Phys. Comm.*, 59, 119
- Nordlund Å., Stein R. F. 1991, in L. Crivellari, I. Hubeny, D. G. Hummer (eds.), *Stellar Atmospheres: Beyond Classical Models*, NATO ASI Series C 341, Kluwer, Dordrecht, p. 263
- Rammacher W., Ulmschneider P. 1992, *A&A*, 253, 586
- Rast M. P. 1995, *ApJ*, 443, 863
- Rast M. P. 1997, in J. C.-D. F.P. Pijpers, C. Rosenthal (eds.), *Solar Convection and Oscillations and their Relationship*, Kluwer Academic Publishers, Dordrecht, p. 135
- Rast M. P., Bogdan T. J. 1998, *ApJ*, 496, 527
- Rast M. P., Nordlund A., Stein R. F., Toomre J. 1993, *ApJ*, 408, L53
- Restaino S. R., Stebbins R. T., Goode P. R. 1993, *ApJ*, 408, L57
- Rimmele T. R., Goode P. R., Harold E., Stebbins R. T. 1995, *ApJ*, 444, L119
- Rossi P., Kalkofen W., Bodo G., Massaglia S. 1992, in M. Giampapa, J. M. Bookbinder (eds.), *Cool Stars, Stellar Systems and the Sun*, Proc. Seventh Cambridge Workshop, Astron. Soc. Pac. Conf. Series, 26, p. 546
- Rutten R. J., Uitenbroek H. 1991, *Solar Phys.*, 134, 15
- Skartlien R. 1998, *A&A*, submitted
- Stein R. F. 1967, *Solar Phys.*, 2, 285
- Stein R. F., Nordlund Å. 1989, *ApJ*, 342, L95

Appendix: Calculating a coarse wavelet transform using the Hilbert- and Fourier transforms

The analytic signal is composed of real and imaginary parts:

$$y_a(t) = y_r(t) + iy_{\text{im}}(t) = A(t) e^{i\Phi(t)},$$

where $y_r(t)$ is the measured real signal, $A(t)$ is the envelope and $\Phi(t)$ is the phase. The instantaneous angular frequency ω is defined by:

$$\omega(t) = \frac{\partial \Phi(t)}{\partial t}.$$

The imaginary part $y_{\text{im}}(t)$ is related to the real part $y_r(t)$ by the Hilbert transform \mathbf{H} :

$$y_{\text{im}}(t) = -\mathbf{H}[y_r(t)] = -\left\{ \frac{-1}{\pi t} * y_r(t) \right\},$$

where the curly braces enclose the Hilbert transform (shown by a convolution in time domain). The full analytic signal is found easily by using the Fourier transform (denoted by tilde):

$$\tilde{y}_a(\omega) = \tilde{y}_r(\omega) - i \{i \operatorname{sgn}(\omega) \tilde{y}_r(\omega)\} = 2H(\omega) \tilde{y}_r(\omega),$$

where H is the Heaviside step function and sgn is the sign function. The curly braces denote the Fourier transform of the Hilbert transform. Implementation is done by standard Fourier transformation (FFT) of the measured signal, and erasing the components at negative frequencies and doubling the remaining components at positive frequencies (last term in the last equation). The inverse transform then yields the analytic signal. Alternatively, multiplying the Fourier transform of the measured signal by $-i \operatorname{sgn}(\omega)$ (as indicated in the last equation) and taking the inverse produces the imaginary component of the analytic signal (Hilbert transform only).

As an example, we wish to extract the envelope, phase and frequency from an AM (amplitude modulated) radio wave $g_r(t)$ with carrier frequency ω_c and audio signal $f(t)$:

$$g_r(t) = f(t) \cos(\omega_c t).$$

The Fourier transform of the wave is:

$$\tilde{g}_r(\omega) = \frac{1}{2} \tilde{f}(\omega + \omega_c) + \frac{1}{2} \tilde{f}(\omega - \omega_c),$$

which is the full bandwidth of $f(t)$ shifted to $\pm\omega_c$ about the origin. The Fourier transform of the imaginary component of the analytical signal is:

$$\tilde{g}_{\text{im}}(\omega) = -i \operatorname{sgn}(\omega) \tilde{g}_r(\omega) = \frac{i}{2} \tilde{f}(\omega + \omega_c) - \frac{i}{2} \tilde{f}(\omega - \omega_c),$$

which has the inverse transform:

$$g_{\text{im}}(t) = f(t) \sin(\omega_c t).$$

The analytic signal is therefore:

$$g_a(t) = f(t) \cos(\omega_c t) + i f(t) \sin(\omega_c t) = A(t) e^{i\Phi(t)},$$

which has the envelope $A(t) = f(t)$, phase $\Phi(t) = \omega_c t$ and instantaneous frequency $\partial \Phi(t) / \partial t = \omega_c$.

This example shows that the bandwidth BW of the envelope must be smaller than the carrier frequency ω_c such that $\omega_c - BW > 0$. For numerical purposes using the discrete Fourier transform, we must also have that $\omega_c + BW < \omega_N$ to avoid aliasing beyond the Nyquist frequency ω_N .

For the concept of envelope and carrier wave to be meaningful, we must also require that several oscillation periods are covered by the envelope, which means that $BW \ll \omega_c$, i.e. that the envelope must be band limited even more than the constraint given above. In general, we deal with broad banded signals that do not satisfy this constraint. We therefore have to study only a limited set of neighboring Fourier components, one at a time, by using bandpass filtered signals. Only then is it meaningful to decompose a signal in envelope and phase.

We therefore apply a bandpass filter (symmetrical cosine filter) on the measured signal, centered on Fourier component ω_c . The resulting bandwidth of the envelope is determined by the width of this filter. We choose the filter width as $W = 2BW = 2\omega_c / N_c$ where N_c is the number of periods of ω_c we wish to have inside the envelope.

The filtered signal does not necessarily correspond to one carrier frequency ω_c plus envelope as in a pure amplitude modulated signal. This will only be the case if the Fourier transform is complex conjugated around ω_c . Deviations from this symmetry introduce a time varying instantaneous frequency that fluctuates around ω_c . This corresponds to an amplitude and phase modulated signal.

If a set of envelopes are calculated, which correspond to different central frequencies of the bandpass, we obtain a coarse wavelet transform (as a function of frequency and time).

... og det var det

