1

A selective overview

JØRGEN CHRISTENSEN-DALSGAARD

Teoretisk Astrofysik Center, Danmarks Grundforskningsfond, and Institut for Fysik og Astronomi, Aarhus Universitet, DK-8000 Aarhus C, Denmark

MICHAEL J. THOMPSON

Space & Atmospheric Physics, The Blackett Laboratory, Imperial College, London SW7 2BZ, UK

1.1 Introduction

Although sometimes ignored, there is no doubt that hydrodynamical processes play a central role in virtually all areas of astrophysics. If they are neglected in the analyses of observations and the modelling, the results for any object must become questionable; the same is therefore true of the understanding of basic astrophysical phenomena and processes that result from such investigations.

Investigations of astrophysical fluid dynamics are hampered by both theoretical and observational problems. On the theoretical side it is evident that the systems being studied are so complex that realistic analytical investigations are not possible. Furthermore, the range of scale, extending in the case of stars from the stellar radius to scales of order 100 m or less, entirely prevents a complete numerical solution. Observationally, the difficulty is to find data that are sensitive to the relevant processes, without being overwhelmed by other, similarly uncertain, effects. Progress in this field therefore requires a combination of physical intuition combined with analysis of simple model systems, possibly also experiments analogous to astrophysical systems, detailed numerical simulations to the extent that they are feasible, together with a judicious choice of observations and development and application of analysis techniques that can isolate the relevant features. Douglas Gough has excelled in all these areas.

In this brief introduction we make no pretense of reviewing the whole vast field of hydrodynamical processes in astrophysics, or even in stars. The excellent contributions to the rest to the book will do a far better job than we can here of discussing the current state and outstanding issues of many aspects of the subject. Nor do we try to review all the many contributions that Douglas has thus far made to the subject. We must content ourselves with a highly subjective selection of a few of the major themes of Douglas's work to date, on this the occasion of his sixtieth birthday, providing some (though again by no means comprehensive) context of associated work in those areas.

At this point two investigations somewhat outside even the broad general range of Douglas's research deserve to be mentioned. One (Gough & Lynden-Bell 1968) was a simple, but ingenious, experiment to study effects of turbulence in a rotating fluid and involving a rather unusual application of Alka-Seltzer tablets. The second (Bastin & Gough 1969), published a few months before the first manned lunar landing, was a computation of the thermal and radiative properties of the lunar surface, as determined by its scales of roughness, and a comparison with the observed thermal properties of the Moon. The resulting inferences of the properties of roughness can surely be characterized as a selenological inverse problem.

1.2 On taking mixing-length theory seriously

In stellar astrophysics, the most obvious hydrodynamical problem concerns convection, the effects of which are directly observable on the solar surface in the granulation. The conditions under which instability arises, *viz.* a density gradient that decreases too slowly with distance from the centre of the star such that an adiabatically rising element of gas finds itself lighter than the surroundings, are well understood, although, as discussed by Gough & Tayler (1966) additional effects such as magnetic fields may substantially complicate the stability analysis. The subsequent development of the instability, on the other hand, and the resulting energy transport and hence the relation between the temperature gradient and the flow of energy in the star, is very uncertain. The 'classic' treatment of convection in stellar modelling is through the so-called mixing-length theory, whereby convection is described by the motion of convective elements over a certain characteristic length, often taken to be a multiple of the local pressure scale height, after which the element is dissolved, delivering its excess heat to the surroundings (*e.g.* Böhm-Vitense 1958).

Analyses of convection often explicitly or implicitly make the *Boussinesque approximation*, where density variations are neglected except in the buoyancy term, the fluid being otherwise treated as incompressible. However, unlike most laboratory experiments, stellar convection typically takes place over regions of very substantial variation in density and hence inherently involves compressibility. The resulting presence of sound waves is a major complication from a numerical point of view, requiring much shorter time steps than those needed to resolve the convective motions. To bypass such complications, Gough (1969) developed the *anelastic approximation* through a formal scale analysis of the fully compressible equations; in this approximation, sound waves are precluded, resulting in a set of equations that are appropriate for the treatment of convection or, *e.g.*, internal gravity waves. As discussed in this volume by Toomre and Elliott the anelastic approximation is still commonly used in large-scale simulations of solar convection.

The visual appearance of laboratory or solar convection in the form of a more or less regular flow pattern suggests a possibly manageable type of numerical computation, whereby the horizontal properties of convection are modelled in terms of an expansion in planforms, perhaps limited to a single term, whereas the vertical behaviour is computed in detail and with substantial numerical resolution. Although still highly simplified, one may hope that such a description can provide physical insight into the behaviour of convection under more realistic circumstances. The basic properties of such *modal descriptions* were elucidated by Gough, Spiegel & Toomre (1975), who included a discussion of the asymptotic limits of weak and strong convective instability, for the case of simple laboratory situations. Numerical solutions for this case were presented by Toomre, Gough & Spiegel (1977). Latour *et al.* (1976) extended the formalism for application to stellar modelling, in the anelastic approximation. This was used by Toomre *et al.* (1976) in the study of the near-surface convection zones in A-type stars, again restricting the description to a single horizontal planform; the results suggested substantial overshoot between the separate hydrogen and helium convection zones found in such stars, suggesting that the intervening region may be mixed, with significant effects on the surface abundances of these stars.†

Even simplified numerical calculations of convection are too complex and time consuming to be included in the general modelling of stars or their pulsations. Thus simpler prescriptions, such as mixing-length theory, are unavoidable. Although mixing-length theory was originally formulated in a heuristic fashion, Spiegel (1963) and Gough & Spiegel (1977) pointed out that it could be derived in terms of a more precise physical picture. This involves a probabilistic description of the creation, motion and destruction of convective elements, depending on the degree of linear instability of the convective motion, and the assumed aspect ratio of the convective elements. For a static star the result, with proper choices of parameters, yielded the

† Interestingly, modern large-scale simulations often use expansions in horizontal planforms, such as in the case of the anelastic spherical-harmonic code discussed by Toomre (this volume). In such simulations, obviously, a very large number of horizontal modes are included.

same result as the normal mixing-length description. However, as developed by Gough (1965), and in more detail by Gough (1977a), the physical description lent itself to generalization to convection in a pulsating star, providing expressions for the perturbations to the convective flux and the turbulent pressure. Baker & Gough (1979) applied the resulting expressions to the modelling of RR Lyrae variables; they showed that convective effects caused a return to stability at sufficiently low effective temperature, providing a natural explanation for the red edge of the instability strip. Spiegel (1963) and Gough (1977b) noted that the treatment of mixing-length theory could be generalized to account for non-local effects, involving both the motion of a given convective eddy over a finite distance of varying stellar conditions and the average over convective eddies yielding the energy transport at a given location. Balmforth & Gough (1990) extended the non-local treatment to pulsating stars. This was applied by Balmforth (1992) to demonstrate that solar acoustic modes are likely stable, the dominant stabilizing effect being the perturbation to the turbulent pressure. Also, Houdek (2000) showed that this could account for the red edge of the δ Scuti and Cepheid instability strip.

1.3 The solar spoon

A strong indication that normal models of stellar evolution might be inadequate resulted from the initial attempts to detect neutrinos from the Sun: the observed upper limit (Davis, Harmer & Hoffman, 1968) was substantially below the model predictions (*e.g.* Bahcall, Bahcall & Shaviv, 1968). Given the simplifications made in modelling the Sun, this was perhaps not surprising. In particular, it was pointed out by, for example, Ezer & Cameron (1968) that mixing of the solar core might reduce the neutrino flux. Such mixing could result from hitherto neglected instabilities of the solar interior. Dilke & Gough (1972) found, based on a simple model of solar oscillations, that g modes might be destabilized by nuclear reactions. They speculated that such instability might trigger sudden mixing of the solar core, through the onset of convective instability. After such a mixing episode the flux of solar neutrinos would be temporarily depressed, over a few million years, and the suggestion was therefore that we are currently in such a period. Dilke & Gough also pointed out that the accompanying relatively rapid fluctuations in the solar luminosity might have acted as a trigger for the series of ice ages that is currently affecting the climate of the Earth, thus making more likely the required special nature of the present epoch. More detailed calculations by Christensen-Dalsgaard, Dilke & Gough (1974), Boury *et al.*

(1975) and Shibahashi, Osaki & Unno (1975) confirmed the reality of the instability, resulting from the build-up of the gradient of the ^{3}He abundance, the first instability setting in after about 200 Myr of evolution. However, the subsequent non-linear development into convective mixing has not been demonstrated. Dziembowski (1983) showed that resonant coupling would limit the amplitudes at a level far below what might be expected to initiate mixing. Furthermore, analyses by Merryfield, Toomre & Gough (1990, 1991), albeit for a simplified physical system, found a tendency for non-linear development to lead to sustained finite-amplitude oscillations rather than direct convective mixing. Even so, it should be kept in mind that solar models are subject to instabilities which have not been properly taken into account in solar modelling.

The closely related issue of the effects of a sustained flow on the solar ^{3}He profile and neutrino flux is addressed by Jordinson (this volume).

As discussed by Shibahashi (this volume), inferences from helioseismology place strong constraints on solar structure; the inferred helioseismic model is in fact very close to the 'standard solar models', including also the predicted neutrino flux. In particular, the excellent agreement between normal, unmixed, models and helioseismic inferences of the solar core indicates that no substantial mixing has taken place, in contrast to the proposal by Dilke & Gough (1972). Although helioseismology does not provide direct information about the solar core temperature upon which the neutrino production predominantly depends, these results nevertheless make it plausible that the cause for the neutrino discrepancy lies in the properties of the neutrino, rather than in errors in solar models. Specifically, if neutrinos have finite mass, the electron neutrinos generated in the nuclear reactions in the Sun may in part be converted to muon or tau neutrinos before reaching the detectors. Strong indications of this process were obtained by measurements from the Sudbury Neutrino Observatory (SNO) which allowed the total neutrino capture rate to be compared with the capture rate of electron neutrinos (Ahmad *et al.* 2001). Direct measurements of all types of neutrinos by Ahmad *et al.* (2002) have very recently provided dramatic confirmation of the transformation between the different types of neutrinos; the total neutrino flux was found to be fully consistent with the predictions of standard solar models, confirming the indications from helioseismology that such models are in fact good representations of solar structure, as reflected in the neutrino production.

It may be somewhat disappointing to Douglas that the conclusion of the efforts, to which he has contributed substantially, over more than three decades to understand the origin of the solar neutrino problem does not point towards novel aspects of stellar interior physics. However, the prospects now of using the predicted solar neutrino production, constrained by the results of helioseismology, to investigate the properties of the neutrinos are perhaps even more exiting.

1.4 Deep roots of solar cycles

The solar 11-year sunspot cycle, and the 22-year magnetic cycle, are clear indications of large-scale dynamical processes in the Sun. These are normally assumed to result from 'dynamo processes', involving a coupling between the solar differential rotation, convection and the magnetic field. However, although models exist which reproduce aspects of the solar cycle, these cannot be regarded as sufficiently definitive that other mechanisms, perhaps involving magnetic oscillations of the deep solar interior, can be ruled out. Indeed, Dicke (1970, 1978) proposed that the periodicity might arise from a regular oscillator in the deep solar interior, modulated by the convection zone to give rise to the apparently somewhat irregular periodicity observed in the solar cycle; he noted that this would be reflected in the long-term phase stability of the cycle, in contrast to the dynamo models where a more erratic phase behaviour might be expected. Gough (1978a, 1981a) analysed examples of the two models, as well as the available data on sunspot maxima and minima; the results suggested that the solar behaviour was intermediate between the models, with no firm conclusion possible.

The variations in solar irradiance accompanying the solar cycle are well established (*e.g.* Willson & Hudson, 1991; Pap & Fröhlich, 1999); there have also been studies, although somewhat conflicting, concerning possible variations in the solar radius (*e.g.* Gavryusev *et al.* 1994; Noël 1997; Brown & Christensen-Dalsgaard 1998). As discussed by Gough (1981a), the changes in solar structure associated with the solar cycle depend on the physical location of the dominant mechanisms causing these changes, thus providing the possibility of obtaining information about the physical nature of the cycle. In particular, he noted that the ratio between the luminosity variations, assumed to be reflected in the irradiance, and the radius variations would depend on the location of the physical mechanism responsible for the solar cycle. Gough analysed the ratio $W = \Delta \ln R / \Delta \ln L$, R being the surface radius and L the luminosity, showing that it increases with increasing depth of the perturbation that produces the variations (see also Däppen 1983). While the irradiance variations, corresponding to an increase of about 0.1 % at sunspot maximum relative to sunspot minimum, has been measured precisely from space, results on the radius variations are somewhat contradictory. Major difficulties are the correction for changes in solar atmospheric structure and, for the visual observations which span the most extended period, corrections for observational bias. It has been pointed out (*e.g.* Parkinson, Morrison & Stephenson 1980) that the width of the path of totality at a total solar eclipse provides an accurate measure of the solar radius, given that the lunar diameter and the ephemeris are known with great accuracy. To apply this idea Gough & Parkinson (1983) carried out determinations of the edges of the eclipse path during the 1983 solar eclipse in Indonesia, by distributing teams of observers to span the northern and southern limits of totality. Unfortunately, some confusion amongst the observers led to inconsistencies in the results which made the desired precise determination impossible.

1.5 Helioseismology: oscillations as a diagnostic of the solar interior

A major development in observational stellar astrophysics took place in 1975 with the first announcements of observations of coherent solar oscillations. Oscillatory signals on the solar surface had been detected substantially earlier, the first detailed results having been obtained by Leighton, Noyes & Simon (1962). These observations showed localized oscillations with periods near five minutes, but with apparently limited spatial and temporal coherence. Thus the oscillations were generally regarded as an atmospheric phenomenon, although a more global nature was suggested by Ulrich (1970), Leibacher & Stein (1971) and Wolff (1972), based on somewhat more detailed observations by Frazier (1968). The global nature of the oscillations was definitely established by the observations by Deubner (1975) and Rhodes, Ulrich & Simon (1977): by obtaining fairly extensive data as a function of both position and time they were able to construct two-dimensional power spectra, as functions of horizontal wave number and frequency, and demonstrate that power was concentrated in ridges corresponding to the modal nature of the oscillations, as earlier predicted by the theoretical analyses. Furthermore, Hill, Stebbins & Brown (1976) announced the detection of oscillations, with periods between 13 and 50 minutes in the apparent solar diameter. Also, Brookes, Isaak & van der Raay (1976) and Severny, Kotov & Tsap (1976) independently presented evidence for an oscillation with a period of very close to 160 minutes in solar Doppler observations.

It was immediately obvious that frequencies of global solar oscillation would be extremely powerful probes of the solar interior. This led to expectations that a 'heliological inverse problem' might be formulated, analogous to

Cambridge University Press
978-0-521-81809-4 - Stellar Astrophysical Fluid Dynamics
Edited by Michael J. Thompson and Jørgen Christensen-Dalsgaard
Excerpt
More information

the inverse problems used in geophysics to infer the structure of the Earth's interior (Christensen-Dalsgaard & Gough 1976). An initial comparison was also made by Christensen-Dalsgaard & Gough between the frequencies reported from the observations by Hill *et al.* (1976) and a solar model; some significance was attached, a little prematurely, by the junior author to the apparently reasonable agreement between observations and model. Indeed, initially, much interest centred on the observations by Hill *et al.* which, since they involved a full solar diameter, appeared to be truly global; in contrast, the modes detected by Deubner and by Rhodes *et al.* were concentrated in a relatively superficial part of the Sun. Thus several additional comparisons were made between the observed frequencies and those of solar models (*e.g.* Scuflaire *et al.* 1975; Iben & Mahaffy 1976; Rouse 1977). Even so, possible alternative explanations unrelated to global solar oscillations were also sought; an interesting hydrodynamical possibility was seiches in supergranules, similar to those observed *e.g.* in a glass of beer (Gough, Pringle & Spiegel, 1976). The longer-period oscillation discovered by Brookes *et al.* and Severny *et al.* was potentially of even greater interest: if it were truly a global solar oscillation it would have to be a g mode, as demonstrated by Christensen-Dalsgaard, Cooper & Gough (1983), and hence it would be sensitive to the deep solar interior. However, the rather peculiar nature of the mode, particularly the long period and the absence of neighbouring modes in the expected dense spectrum of g modes, prompted alternative explanations; in particular, the close proximity of the period to 1/9 of a day led to suspicions that the observations might be related to phenomena in the Earth's atmosphere.

Although of a more superficial nature, the original high-degree five-minute oscillations are of very substantial diagnostic potential for the solar interior. They are concentrated in the upper parts of the convection zone. However, since the convection zone is essentially adiabatically stratified, apart from a very thin upper layer, its structure is fully determined by specifying the (constant) value of the specific entropy. As a result, the constraints from the five-minute oscillations on the upper parts of the convection zone essentially allow extrapolation to the rest of the convection zone. Consequently, Gough (1977c) and Ulrich & Rhodes (1977) were able, on the basis of the early data, to infer that the convection zone was substantially deeper than obtained from the then current solar models. Also, the modes are directly sensitive to the equation of state in the ionization zones of hydrogen and helium, providing a first helioseismic test of the physics of the solar interior (Berthomieu *et al.* 1980; Lubow *et al.* 1980). Gough (1984a) emphasized the great significance of using such analyses to test the highly complex thermodynamic properties of matter under solar conditions, as well as to determine the present solar envelope helium abundance. As discussed by Däppen in this volume, such investigations of the 'microphysics' of the solar interior have been very successful, allowing tests of subtle aspects of the equation of state in the solar convection zone.

The oscillations are also sensitive to motions in the solar interior. Deubner, Ulrich & Rhodes (1979) noted that the frequencies of the five-minute oscillations were shifted by the sub-surface rotational velocity. More general rotation laws were considered by Gough (1981b), and independently by Hansen, Cox & van Horn (1977) and Cuypers (1980), thereby laying the foundation for the use of observed rotational frequency splittings to investigate solar internal rotation, by extending earlier work by Ledoux (1951) who had considered uniform rotation.

It is interesting that much of the early enthusiasm for helioseismology was inspired by the diameter observations by Hill *et al.* and the detection of the 160-minute oscillation. Later observations, including the detailed data obtained for periods below 15 minutes and the extensive attempts to detect long-period oscillations, make it overwhelmingly likely that the early results were in fact unrelated to global solar effects. However, they served an important role as inspirations for other observational efforts which have led to the dramatic success of helioseismology based on the five-minute oscillations, also evident in the present volume.

1.6 Inverting helioseismic data

A key issue in the applications of helioseismology is evidently the ability to carry out *inversion*, *i.e.*, to infer from the observed frequencies localized information about the properties of the solar interior. Similar problems have a long history of study in geophysics.† It was realized by Gough (1978b) that the method proposed by Backus & Gilbert (1968) was well suited to obtain inferences of, for example, solar rotation in the form localized averages characterized by well-defined *averaging kernels*. Establishing a procedure that has been used extensively since, Gough tested the method on artificial data similar to the observations presented by Deubner *et al.* (1979). Gough (1982) analyzed rotational splittings inferred from observations of the solar diameter by Bos & Hill (1983), to constrain the solar internal rotation and thence the gravitational quadrupole moment J_2 of the Sun, which is of obvious importance to tests based on planetary motion of gravitational

† A discussion of the information transfer between geo- and helioseismology was provided by Gough (1996a).

theories. In addition to the Backus-Gilbert technique he applied a least-squares polynomial fit as well as a minimization of the integral of the square of the radial gradient of angular velocity, subject to the constraints of the observed splittings being satisfied exactly. It is interesting to note that the latter technique resembles the regularized least-squares fitting technique now commonly used in helioseismic analyses.

The solar quadrupole moment affects the gravitational field outside the Sun and hence the motion of planets. This effect must be taken into account in tests of general relativity based on, for example, the precession of the orbit of Mercury. Given the value of J_2 obtained by Gough (1982), radar observations of planetary motion were indeed consistent with general relativity. Although, as already mentioned, the observations of oscillations in the solar diameter were questionable, more recent inferences of J_2 based on helioseismic inversions of frequencies in the five-minute region have confirmed this conclusion (*e.g.* Pijpers 1998; Roxburgh 2001).

Douglas, with his student A. Cooper, also developed inversion techniques for the inference of solar internal structure. Since the dependence of the oscillation frequencies on, *e.g.*, the sound speed and density of the solar interior is strongly nonlinear, this is normally carried out by linearization around a reference model, relating the frequency differences between observations and model to the differences in structure between the Sun and the model. By means of tests involving structure and frequency differences between simplified solar models, they investigated the ability of the Backus-Gilbert technique to construct localized information about the density difference between the Sun and the model. The success depended greatly on the assumed mode set (Cooper 1981; Gough 1984b): with just low-degree p modes only broad averages in the outer parts of the model could be obtained, whereas with a set including also low-degree g modes good resolution was possible throughout the model. These results were of substantial significance in the evaluation of the DISCO mission to observe solar oscillations, proposed to ESA (*e.g.* Bonnet *et al.* 1981; Balogh *et al.* 1981): it was estimated that this mission would indeed be able to observe low-order, low-degree g modes. While DISCO was in the end not selected, and the detection of g modes, even from the much more powerful instruments on the SOHO spacecraft, remains elusive (*e.g.* Appourchaux *et al.* 2000), the detailed p-mode data have dramatically realized the potential for structure inversion (Shibahashi, this volume).

Approximate analyses of stellar oscillation frequencies have played a major role in the development of helio- and asteroseismology. An early example was the derivation by Gough, Ostriker & Stobie (1965) of approximate expres-