We propose that the kind of stellar variability exhibited by the sun in its magnetic activity cycle should be considered as a prototype of a class of stellar variability. The signature includes long 'periods' (compared to that of the radial fundamental mode), erratic behavior, and intermittency. As other phenomena in the same variability class we nominate the luminosity fluctuations of ZZ Ceti stars and the solar 160^m oscillation. We discuss the possibility that analogous physical mechanisms are at work in all these cases, namely instabilities driven in a thin layer. These instabilities should be favorable to grave modes (in angle) and should arise in conditions that may allow more than one kind of instability to occur at once. The interaction of these competing instabilities produces complicated temporal variations. Given suitable idealizations, we show how to begin to compute solutions of small, but finite, amplitude and we discuss the prospects for further developments.

THE PROPOSAL

VARIABLE VARIABILITY

An aim of this paper is to argue that the kind of variability that the sun displays in its magnetic activity cycle is the prototype for a category of stellar variability that should be isolated and studied as a generic phenomenon. We are not referring here to the group of solar type stars that show magnetic activity, though they and the sun do make up a class of variable star in the usual sense of the term. We are speaking of a kind of variability and thus of a broader category, if a more abstract one, than that of a class of star. The kind of variability that we have in mind includes intermittency, such as the sun exhibited

* The financial support of this work by the National Science Foundation under Grants NSF PHY-77-27086 at Columbia and NSF MCS-79-02766 is gratefully acknowledged.

when it went through the Maunder minimum [1], and irregularity such as the cycle displays when it is detectable. We presume that all solar type variables show this variable variability, as we may call it. We suggest that the time dependence of ZZ Ceti stars [2] is of the same kind, details aside. The stars in this latter class are variable white dwarfs with marked temporal intermittency and, if they are in the same variability category as the sun, this could be useful for solar studies since ZZ Ceti periods are of the order of ten minutes. We believe too that the sun shows variable variability in several ways and in particular through the 160^m oscillation, which is sometimes quite hard to detect and is always very noisy [3]. There are other examples that come readily to mind, but we mention only these since they are the only ones that we have done serious calculations for in the way that we shall describe below.

The point of isolating a kind of variability is that it may help to identify the physical mechanism that produces the variations. In the case of the simplest kind of stellar variability, that exemplified by the regularity of Cepheids, we normally try to find a mechanism analagous to that of the Cepheids, in which case we might expect that a static star is overstable or vibrationally unstable to small perturbations. Then periodic solutions will bifurcate from the static solutions in a fashion called a Hopf bifurcation [4] nowadays. Overstability is more complicated than ordinary or direct or dynamical instability, in which perturbations to a static configuration grow monotonically and steady solutions bifurcate from the static ones.

MODELS OF APERIODICITY

The contrast between direct instability and overstability is made vivid by systems that can manifest either depending on the value of some system parameter such as angular velocity. We suggest that such a system is implicated in the solar cycle. Qualitative evidence for this remark is provided by a model constructed to explain the solar 5^m oscillation as a convective overstability [5]. The model could be either overstable or directly unstable according to the values of certain parameters. As in most such systems there are two important parameters, one controlling the amount of each instability. The system is said to have codimension two [6]. Such a system can be rewired, so to say, so that one parameter controls the relative amounts of both kinds of instability and another controls the total amount of instability. When the system is set to hover between the two instabilities and the amount of instability is turned up, the resulting oscillations become aperiodic. This led to the conjecture that unstable systems hovering between the two kinds of instability would generally behave aperiodically. This suggestion was tested on a one-zone model for radial pulsation [7]. The idea was that for a mean $\Gamma < 4/3$ one gets direct (or dynamical) instability while for $\Gamma > 4/3$ one may have overstability. For $\Gamma \approx 4/3$ one might expect erratic behavior, and that was what was found. This is the kind of behavior we are postulating to explain the time dependence of the solar cycle and of other variable variables.

We have to ask at once whether it is reasonable to expect to find many systems arranged not only to be unstable but also to be unstable in several ways. Our answer is yes but our reasons are complicated. Let us simply say here that one reason lies in the circumstance that many systems do have long periods compared to their natural radial periods. To see what this may mean, consider the case of convective overstability. How may we engender slow variations in the elementary theory of convection?

CELL SIZES AND TIME SCALES

When you heat a fluid from below to induce convection you generally are doing something thermally complicated, but approximately you are usually fixing the temperatures on the top and bottom boundaries of the fluid. When the imposed temperature difference is large enough, sustained motions may begin [8]. These motions are organized into cells which tesselate the layer on a horizontal scale comparable to the layer thickness. However, if you instead specify the heat flux on the boundaries, the cells are as large in the horizontal as the geometry will allow [9]. Such big cells are easier to excite and are slower to react than the more popularly sized cells.

In a case where overstability is also possible, the role of the boundary conditions is very important too. For, not only are the growth rates small, as in normal convection, but the frequencies of the overstable modes are also small. Hence boundary conditions that favor large horizontal scales in rather thin layers will tend to put such systems willy-nilly into the states we want them. Whether those boundary conditions are realistic in a given configuration cannot be stated in advance, but at least they are frequently not implausible. Moreover, it is not always easy to say ahead of time precisely which conditions will induce large horizontal scales and relatively low frequencies. But that they may arise naturally is attested by studies of instabilities relating to ZZ Ceti stars [10,11]. The periods turn out to be long compared to the fundamental period of the star which is reckoned in seconds. This is just the kind of situation that we need for the analysis we shall describe.

GETTING STARTED

Even if it is true that mild instabilities in thin shells lead to the variable variability we have described, it might not always be obvious which thin shell is involved. However once the idea is there, we have the motivation to look for the right kind of instability. For the solar cycle a clue is provided by the familiar problem [12] of building strong magnetic fields in many solar dynamo models. The general idea is that differential rotation produces a toroidal field whose strength builds till it goes unstable and buckles to protrude from the sun and create spots. The difficulty in this picture is that turbulent convection and magnetic buoyancy [12] will quickly destroy any ordered field so it is not clear how a significant toroidal field can persist for any time in the solar convective zone. A way out of this difficulty is to form the ordered field just below the convective zone [13]. Large scale convection produces flux expulsion [12] and topological pumping [14]. However, since the material just below the convection zone is a good conductor, it does not readily allow the field to penetrate it. The field is nevertheless swept into an intermediate region by penetrative convection, which may be mild enough to leave it ordered. As the field builds up in the bottom of the penetrative zone, the penetrative motion will be impeded and the convection zone recedes leaving behind a layer of ordered field. Ultimately, the layer becomes thick enough for magnetic buoyancy to overwhelm the local stable temperature gradient. There follows a new round of solar activity. Of course other observable manifestations of such a process should exist and it is unclear as yet whether we are on firm grounds [15].

THE 160^m OSCILLATION

For another illustration of the procedure consider the 160^m oscillation. Here there is no accepted explanation and our view is that the nature of the oscillation leads us to look for a thin overstable layer. The low frequency points to gravity waves and perhaps to an important role for buoyancy. The popular objection to explaining the 160^m period as owing to gravity waves is that the spectrum should be dense and some claim that a broad band of frequencies should be excited. However, if we are dealing with a thin layer, we may study waves whose lengths are much greater than the layer thickness. It is then possible to strike a balance between nonlinearity and dispersion so that a solitary wave or a train of them is produced as in the theory of shallow water waves [16]. In this picture, which may be described by our procedures, the 160^m period should be the travel time of the solitary wave around the sun. This gives a clue to where the waves are excited.

In a stably stratified gas, linear gravity waves propagate with a maximum speed given roughly by [17]

$$c_g = \left(\frac{\gamma-1}{\gamma} gH\right)^{\frac{1}{2}} = \left(\frac{\gamma-1}{\gamma} R_* T\right)^{\frac{1}{2}},$$

where R_* is the gas constant. For present purposes we can write this as

$$c_g \sim 10^6 \sqrt{T_*} \text{ cm/s},$$

where $T_4 = T/10^4 K$. If the layer in question has radius $7\times10^{10} r$ cm, we find a period

$$P = 10^4 \, r/\sqrt{T_4} \text{ min.}$$

So we want a layer such that

$$T_4 = 3600r.$$

This layer needs to be in the deep interior. We propose that the argument makes it worthwhile to inquire whether the 160^m oscillation can be excited in the solar core.

Fortunately, a suitable mechanism is already known: Dilke and Gough [18] found an overstability that should exist in the present solar core. For this instability to be in keeping with our prescription we must postulate a thin layer rich in ^3He at the edge of the nuclear burning region. Granted this, the kind of nonlinear analysis that we shall now illustrate goes through. It has been done for a mildly nonBoussinesq model and it needs more refinements. Nevertheless there are some interesting features of the solutions. For example, a mild thermal anomaly propagates about the core. This contributes to energy generation, and to get the right solar luminosity, we have to lower the central temperature a bit. How significant the change is depends on the amount of ^3He we put in; a careful comparison with the oscillation data will be needed to make a quantitative statement. In fact no numerical estimates of any kind are given in the next section, which is too physically bare to be anything but a demonstration model. As we shall explain below, it is too primitive mathematically. Yet we think its general design is good and if you are looking for an approach to these problems, you might want to consider this possibility.

THE PROCEDURE

FORMULATION OF A TRACTABLE PROBLEM

The General Model

To discuss the dynamics of a hypothetical magnetic layer located just under the solar convective zone, we might reasonably presume that this layer is subject to a given heat flux from below and that it passes this same flux on to the layers above it. Similarly, we might guess that the overlying convective zone pumps a horizontal magnetic field downward into the layer. In the static state the field continues to diffuse slowly downward, and there will be a slight time dependence. We may avoid this complication by assuming that the field is removed from the bottom boundary of the layer at the rate that it is being fed in at the top. This is the idealization we propose for the sun's magnetic layer. For semiconvection we would likewise assume that the flux of helium at the top and bottom boundaries is prescribed.

We assume that the geometry is plane-parallel and let z be the upward coordinate, y be the horizontal coordinate corresponding to the eastward direction and x be the equivalent northward direction. The fluid is confined to the layer $-d/2 \leq z \leq d/2$.

Though the calculations that follow are detailed, they are stripped down to the barest essentials for this demonstration. No extras are included. Thus we assume here that all of the parameters characterizing the fluid, such as the gas constant, the acceleration of gravity, the specific heat at constant pressure, the permeability, and the several diffusivities, are constants. The main dependent variables of interest for the convective process are the velocity and the state variables such as temperature, magnetic field and molecular weight. To describe state variables we shall use a standardized notation which we illustrate for the case of temperature.

The Dependent Variables

Let the temperature, T, be decomposed into a static part and a convective part:

$$T(x,z,t) = \bar{T}(z) + \delta T(x,z,t)$$

where t is time. We allow no y-dependence and consider only a two-dimensional problem. Let

$$T_o = \bar{T}(0)$$

be used as a temperature scale and let

$$\xi_T = \frac{d}{T_o}\frac{d\bar{T}}{dz}.$$

This measures the temperature contrast across the static layer. A scaled temperature disturbance is then defined by

$$\delta T = \left| \xi_T + \frac{g\,d}{C_p T_o} \right| T_o \, \theta,$$

where C_p is the ratio of specific heats and g is the acceleration of gravity. For the other state variables we proceed similarly. For example we might introduce a scaled magnetic perturbation in the two dimensional case where the field has only a y-component:

$$\delta B = |\xi_B - \xi_\rho| B_o \Sigma,$$

where ρ is density. An analogous scaled variable would be used for molecular weight.

We shall assume that the velocity is solenoidal, and this is not a bad approximation for the kind of gravity wave solution we are after. The velocity can therefore be described by a non-dimensional stream function:

$$\mathbf{u} = \frac{\kappa}{d}\left(-\frac{\partial\psi}{\partial z}\ ,\ 0,\ \frac{\partial\psi}{\partial x}\right),$$

where κ is the thermal diffusivity.

The Boussinesq Equations

Next we make an approximation which forces us to give up some qualitative features of the problem that are of astrophysical interest, but which permits us to bring out the basic nature of the phenomenon as simply as possible and to illustrate the kind of calculation that the subject entails. This is the Boussinesq approximation in which we omit density fluctuations except insofar as they directly produce driving by buoyancy forces. Boussinesq theory also neglects pressure fluctuations in the equation of state. The pressure fluctuation enters only as a gradient term in the equation of motion as needed to maintain the solenoidal condition on the velocity field. In hydromagnetic convection the analogous approximation is the neglect of fluctuations in the total pressure (gas plus magnetic). On introducing these approximations we are led to the following equations for two-dimensional Boussinesq magnetoconvection [19]:

$$(\partial_t - \sigma\nabla^2)\nabla^2\psi = -\sigma\zeta_\Theta R\partial_x\Theta + \sigma\tau\zeta_\Sigma S\partial_x\Sigma + \frac{\partial(\psi,\nabla^2\psi)}{\partial(x,z)}, \qquad (1)$$

$$(\partial_t - \nabla^2)\Theta - \partial_x\psi = \frac{\partial(\psi,\Theta)}{\partial(x,z)}, \qquad (2)$$

$$(\partial_t - \tau\nabla^2)\Sigma - \partial_x\psi = \frac{\partial(\psi,\Sigma)}{\partial(x,z)}, \qquad (3)$$

where

$$R = \frac{gd^3}{\kappa\nu}\left|\xi_T + \frac{g\,d}{C_p T_o}\right|, \qquad S = \frac{gd^3}{\eta\nu}\,\frac{B_o^2}{\mu p_o}\left|\xi_B - \xi_\rho\right|,$$

$$\sigma = \nu/\kappa\ ,\ \tau = \eta/\kappa\ .$$

Here ζ_F is +1 if the vertical gradient of F is in the unstable sense and is -1 if it is in the stable sense (F may be Θ or Σ), p stands for gas pressure, ν for kinematic viscosity, μ for permeability, and η for magnetic diffusivity. The Rayleigh (R and S) and Prandtl (σ and τ) numbers appear because we have used natural units for length (d) and time (d^2/κ). We have followed one standard practice in assuming that the Rayleigh numbers are positive and separating out the stability discriminants, ζ_F. We find this cumbersome and we prefer to adopt a particular choice for them:

$$\zeta_\Theta = 1, \ \zeta_\Sigma = -1,$$

which is the typical case for semiconvection [20]. We make this choice even though semiconvection plays a secondary role in destabilizing ZZ Ceti stars [10] because it is one of the few cases for which we know explicitly the right discriminants to select. For the magnetoconvective model of the solar cycle, a number of possible ζ combinations may arise in the important magnetic layer and, in any case, for that situation, there appear to be vital nonBoussinesq effects. (We suspect too that differential rotation is important and that adds another ζ.) So it would be misleading to try to describe a solar model at the present minimal level. But elementary as the present example is, it seems to contain the essential nonlinear dynamics of semiconvection, which may play a very significant role in certain variable stars and whose time dependence, we suspect, is a good example on which to base thinking about variable variability.

The Boundary Conditions

When we go from magnetoconvection to semiconvection, in two dimensions, we have to deal with precisely the same set of Boussinesq equations, but now Σ is the perturbation in molecular weight instead of the magnetic disturbance. These equations are also the right ones for thermohaline convection, which has a much greater following [21] and which has inspired our notation (Σ for salinity, S for saline Rayleigh number). All the computations we know of in this subject have been done with Θ and Σ vanishing on the upper and lower boundaries [24]. As advertised, we here fix fluxes on the upper and lower boundaries. Hence, the pertubation fluxes must vanish on these boundaries and we require that

$$\partial_z \Theta = 0, \ \partial_z \Sigma = 0 \quad @ \ z = \pm\tfrac{1}{2}. \qquad (4a,b)$$

We also need kinematic conditions on top and bottom and we assume that the boundaries are stress free but not deformable. Then

$$\psi = 0, \ \partial_z^2 \psi = 0 \quad @ \ z = \pm\tfrac{1}{2}. \qquad (5a,b)$$

An advantage of having the fixed-flux boundary conditions is that they generally favor large horizontal scales and once this is realized, we can use the methods generally associated with the theory of shallow water waves [16,22]. This permits us to include in our asymptotic studies several effects that have been difficult to treat other than by numerical methods. But the main physical interest of these large horizontal scales is that they also involve slow behavior such as we would like to postulate in a model for variable variability.

The Scaled Equations

To find approximate solutions we proceed by asymptotic methods, combining amplitude expansions with scaling of time and space coordinates. To ensure small amplitude, we assume that the layer is only mildly unstable. Let

$$R = R_0 + \varepsilon^2 R_2, \qquad S = S_0 + \varepsilon^2 S_2,$$

where $\varepsilon^2 \ll 1$ and R_0 and S_0 are values of R and S that render the fluid neutrally stable while R_2 and S_2 are arbitrary.

We anticipate that, as in ordinary convection theory, the amplitude of the motion is $O(\varepsilon)$ and we therefore let

$$\psi = \varepsilon \Psi .$$

However it is best not to rescale the temperature in Boussinesq convection with fixed flux [23]. As with many nonlinear problems, we expect a close connection between amplitude and (here spatial) periodicity and we accordingly scale the horizontal coordinate to be proportional to amplitude. We also assume long time scales, but the factor by which we stretch the time is found essentially by trial and error. We let

$$\tilde{x} = \varepsilon x, \qquad \tilde{t} = \varepsilon^3 t.$$

Then we follow a deplorable notational trick that is widely used in fluid dynamics and drop the tildes. The reason is that the tildes make the equations cumbersome to read while the other letters we might have used for a rescaled time are preempted. The equations that we want to study now are

$$\theta_{zz} = \varepsilon^2 (\Psi_x - \theta_{xx} + \Psi_z \theta_x - \Psi_x \theta_z) + \varepsilon^3 \theta_t \tag{6}$$

$$\tau \Sigma_{zz} = \varepsilon^2 (\Psi_x - \tau \Sigma_{xx} + \Psi_z \Sigma_x - \Psi_x \Sigma_z) + \varepsilon^3 \Sigma_t \tag{7}$$

$$\begin{aligned} \Psi_{zzzz} = R\theta_x - \tau S \Sigma_x + \varepsilon^2 [-2\Psi_{xxzz} + \frac{1}{\sigma}(\Psi_z \Psi_{zzx} - \Psi_x \Psi_{zzz})] \\ + \frac{\varepsilon^3}{\sigma}\Psi_{tzz} + \varepsilon^4 [-\Psi_{xxxx} + \frac{1}{\sigma}(\Psi_z \Psi_{xxx} - \Psi_x \Psi_{xxz})] + \frac{\varepsilon^5}{\sigma}\Psi_{txx} . \end{aligned} \tag{8}$$

where the subscripts z,x,t signify partial differentiation with respect to the variables indicated.

Now we set out to find asymptotic solutions for small ε taking the other parameters, σ and τ, to be of order unity. This may not be the best choice for astrophysics, but it is often used by numerical experimenters and it provides a simple introduction.

EXPANSIONS

For small ϵ we consider expansions of the form,

$$\Psi = \Psi_0 + \epsilon\Psi_1 + \epsilon^2\Psi_2 + \ldots ,$$
$$\theta = \theta_0 + \epsilon\theta_1 + \epsilon^2\theta_2 + \ldots ,$$
$$\Sigma = \Sigma_0 + \epsilon\Sigma_1 + \epsilon^2\Sigma_2 + \ldots .$$

The boundary conditions are

$$\theta_{nz}=0,\ \Sigma_{nz}=0,\ \Psi_n=0,\ \Psi_{nzz}=0 \quad @ \quad z=\pm\tfrac{1}{2}.$$

From (6)-(8) we obtain

$$\theta_{0zz} = 0, \quad \Sigma_{0zz} = 0 \qquad (9a,b)$$

and

$$\Psi_{0zzzz} = (R_0 f - \tau S_0 g)_x . \qquad (10)$$

The solutions are

$$\theta_0 = f(x,t), \quad \Sigma_0 = g(x,t) ,$$

where f and g are arbitrary functions to be determined and

$$\Psi_0 = (R_0 f - \tau S_0 g)_x P(z)$$

where

$$P(z) = \frac{1}{4!}(z^4 - \tfrac{3}{2} z^2 + \tfrac{5}{16}).$$

In exactly the same way we find that

$$\theta_1 = f_1(x,T), \quad \Sigma_1 = g_1(x,T)$$

where f_1 and g_1 are also arbitrary so far and

$$\Psi_1 = (R_0 f_1 - \tau S_0 g_1)_x P(z) .$$

Next we find that

$$\theta_{2zz} = \Psi_{0x} - \theta_{0xx} + \Psi_{0z}\theta_{0x} - \Psi_{0x}\theta_{0z}$$

$$\tau\Sigma_{2zz} = \Psi_{0x} - \tau\Sigma_{0xx} + \Psi_{0z}\Sigma_{0x} - \Psi_{0x}\Sigma_{0z}$$

which may be written as

$$\theta_{2zz} = (R_0 f - \tau S_0 g)_{xx} P - f_{xx} + f_x (R_0 f - \tau S_0 g)_x P' , \qquad (11a)$$

$$\tau\Sigma_{2zz} = (R_0 f - \tau S_0 g)_{xx} P - \tau g_{xx} + g_x (R_0 f - \tau S_0 g)_x P' . \qquad (11b)$$

Since (11a,b) are inhomogeneous forms of (9a,b), we know that we have to remove the secular terms. If we integrate (11a,b) in z from $-\frac{1}{2}$ to $+\frac{1}{2}$ we find that the integrals on the left side vanish. Hence, so must the integrals on the right side, and the expression of that fact gives the solvability conditions for (11a,b). On noting that the integral of P is $(5!)^{-1}$, we obtain the conditions

$$\mathbf{A}\begin{Vmatrix} f_{xx} \\ g_{xx} \end{Vmatrix} = 0; \quad \mathbf{A} = \begin{Vmatrix} R_0 - 5! & -\tau S_0 \\ R_0 & -\tau(S_0 + 5!) \end{Vmatrix}. \tag{12}$$

For (12) to have nontrivial solutions we require that $\det \mathbf{A} = 0$, whence we find that

$$R_0 - S_0 = 5!$$

This determines a critical value for the total Rayleigh number, R-S, which is just that found for the Bénard problem with fixed heat flux [9]. We also may note that the right and left null vectors of $\mathbf{A}$ are respectively

$$\mathbf{r} = \begin{Vmatrix} 1 \\ \tau^{-1} \end{Vmatrix}, \qquad \mathbf{l} = \| R_0, \ -S_0 \|.$$

We see that (12) is satisfied for

$$f = \tau g.$$

Now we may solve (11) and we obtain

$$\Theta_2 = f_2(x,T) + H_2(z) f_{xx} + G_2(z)(f_x)^2$$

$$\tau^2 \Sigma_2 \quad \tau^2 g_2(x,t) + \tau H_2(z) f_{xx} + G_2(z)(f_x)^2,$$

where f_2 and g_2 are yet two more functions to be found and

$$G_2 = z^5 - \frac{5}{2} z^3 + \frac{25}{16} z$$

and

$$H_2 = \frac{1}{6}\left(z^6 - \frac{15}{4} z^4 + \frac{27}{16} z^2 - \frac{99}{896}\right),$$

where an arbitrary constant has been chosen for later convenience.

Next from (8) we find, after some reductions, an equation for $\Psi_{2_{zzzz}}$. This is easily solved and we obtain

$$\Psi_2 = [R_0 f_2 - \tau S_0 g_2 + (R_2 - S_2) f]_x P(z) + P_2(z) f_{xxx} + Q_2(z)_x f_{xxx}$$

where

$$P_2 = \frac{(5!)^2}{10!}\left(z^{10} - \frac{45}{4} z^8 - \frac{483}{8} z^6 + \frac{18675}{64} z^4 - \frac{96345}{256} z^2 + \frac{9835}{128}\right).$$

Q_2 is an odd polynomial, which is all we need to know about it.

And so it goes. At ε^3 the time derivatives appear:

$$\Theta_{3_{zz}} = \Psi_{1_x} - \Theta_{1_{xx}} + \Psi_{1_z}\Theta_{0_x} + \Psi_{0_z}\Theta_{1_x} + \Theta_{0_t}$$

$$\tau\Sigma_{3_{zz}} = \Psi_{1_x} - \tau\Sigma_{1_{xx}} + \Psi_{1_z}\Sigma_{0_x} + \Psi_{0_z}\Sigma_{1_x} + \Sigma_{0_t}$$

We integrate these two equations in z across the layer as we did in the previous orders. We obtain a pair of differential equations, or a matrix differential equ tion. We multiply this by **1**, the left null vector of **A**, and obtain the solvability condition

$$(R_0 - S_0/\tau)\, f_t = 0.$$

This means that either f_t or its coefficient must vanish. If the former is true, we have chosen the wrong time scale. Indeed that is so if we want certain kinds of solutions. For example, if $\tau=1$ or if both ζ_Θ and ζ_Σ are positive, we expect steady convection and should scale the time accordingly. However, for the problem we are studying, we choose to let

$$R_0 - S_0/\tau = 0.$$

Then we have

$$R_0 = \left(\frac{1}{1-\tau}\right)5!, \quad S_0 = \left(\frac{\tau}{1-\tau}\right)5! \qquad (13a,b)$$

Having thus removed secular terms we may solve the two equations that result from the z-integration and we get

$$f_{1_{xx}} - \tau g_{1_{xx}} = -\frac{1-\tau}{\tau} f_t . \qquad (14)$$

We are now near the end and may at last skip to order ε^4, which leads to the equations

$$\Theta_{4_{zz}} = \Psi_{2_x} - \Theta_{2_{xx}} + \Psi_{0_z}\Theta_{2_x} - \Psi_{0_x}\Theta_{2_z} + \Psi_{1_z}\Theta_{1_x} + \Psi_{2_z}\Theta_{0_x} + \Theta_{1_t} \qquad (15a)$$

$$\tau\Sigma_{4_{zz}} = \Psi_{2_x} - \tau\Sigma_{2_{xx}} + \Psi_{0_z}\Sigma_{2_x} - \Psi_0\Sigma_{2_z} + \Psi_{1_z}\Sigma_{1_x} + \Psi_{2_z}\Sigma_{0_x} + \Sigma_{1_t} . \qquad (15b)$$

Now we integrate (15a,b) over z and we find that

$$f_{1_t} + \int_{-1/2}^{+1/2} [\Psi_{2_x} - \Theta_{2_{xx}} + \Psi_{0_z}\Theta_{2_x} - \Psi_{0_x}\Theta_{2_z}]\, dz = 0 \qquad (16a)$$

and

$$g_{1_t} - \int_{-1/2}^{+1/2} [-\Psi_{2_x} + \tau\Sigma_{2_{xx}} - \Psi_{0_z}\Sigma_{2_x} + \Psi_{0_x}\Sigma_{2_z}]\, dz = 0. \qquad (16b)$$

It is then straightforward to derive from (14) and (16a,b) an equation for f. The final step involves some integrations and we obtain then the evolution equation

$$f_{tt} - \mu\tau f_{xxxx} - \kappa\tau f_{xxxxxx} - \nu[(f_x)^3]_{xxx} = 0, \qquad (17)$$

where

$$\mu = (R_2 - S_2)/5!,$$

$$\int_{-1/2}^{+1/2} (P_2-H_2)\,dz = .1967893\ldots \equiv \kappa,$$

and

$$(5!)^2 \int_{-1/2}^{1/2} P^2 dz = 1.2301587\ldots \equiv \nu.$$

Equation (17) is a nonlinear wave equation whose properties we are attempting to understand. Here we sketch one of its approximate solutions that gives the flavor of the answers we seek.

BUOYANCY WAVES

When the amplitude of f is infinitesimal, the evolution equation may be linearized and it has a solution of the form

$$f = e^{\eta t} \cos(kx).$$

This gives us

$$\eta^2 = \tau k^4 (\mu - \kappa k^2).$$

So we have instability whenever

$$\mu \geq \mu_0 \equiv \kappa k^2.$$

If the situation is only slightly unstable we can once again make an amplitude expansion. Let

$$\mu = \mu_0 + \frac{\lambda}{\tau}\delta^2$$

where $\delta^2 << 1$ and λ is an arbitrary parameter; δ is analogous to ε in the previous development and λ is analogous to $R_2 - S_2$. Hence we scale the amplitude with δ and set

$$f = \delta F.$$

We also define a slow time

$$s = \delta t.$$

The evolution equation becomes

$$\kappa\tau(F_{xxxxxx} + k^2 F_{xxxx}) = \delta^2\{F_{ss} - \lambda F_{xxxx} - \nu[(F_x)^3]_{xxx}\}.$$

We expand again:

$$F = F_0 + \delta F_1 + \delta^2 F_2 + \dots \quad .$$

In leading order we find the linear problem

$$F_{0xxxxxx} + k^2 F_{0xxxx} = 0 \ ,$$

with the solution

$$F_0 = X(s)\cos(kx) + Y(s)\sin(kx) \ ,$$

where X and Y are arbitrary functions. Then in the next order,

$$\kappa\tau(F_{1xxxxxx} + b^2 F_{1xxxx})$$
$$= (\ddot{X} - \lambda k^4 X)\cos(kx) + (\ddot{Y} - \lambda k^4 Y)\sin(kx)$$
$$- \nu k^3 [X\sin(kx9 - Y\cos(kx)]_{xxx} \ .$$

We multiply by sin(kx) and integrate from 0 to $2\pi/k$; then we multiply by cos(kx) and integrate. This leads to coupled equations for X and Y. Rather than write these directly we prefer to use as variables $\mathcal{A}$ and ϕ where

$$X = \mathcal{A}\cos\phi, \quad Y = \mathcal{A}\sin\phi.$$

Then

$$F = \mathcal{A}\cos(kx + \phi).$$

The equations are

$$\ddot{\mathcal{A}} - \mathcal{A}\dot{\phi}^2 - \lambda k^4 \mathcal{A} - \tfrac{3}{4}\nu k^5 \mathcal{A}^3 = 0 .$$

and

$$\mathcal{A}\ddot{\phi}^2 + 2\dot{\mathcal{A}}\,\dot{\phi} \quad 0 .$$

We find that

$$\dot{\phi} = b/\mathcal{A}^2$$

where b is arbitrary and we get the amplitude equation

$$\ddot{\mathcal{A}} - b^2/\mathcal{A}^3 - \lambda k^4 \mathcal{A} - \tfrac{3}{4}\nu k^5 \mathcal{A}^3 = 0 . \tag{18}$$

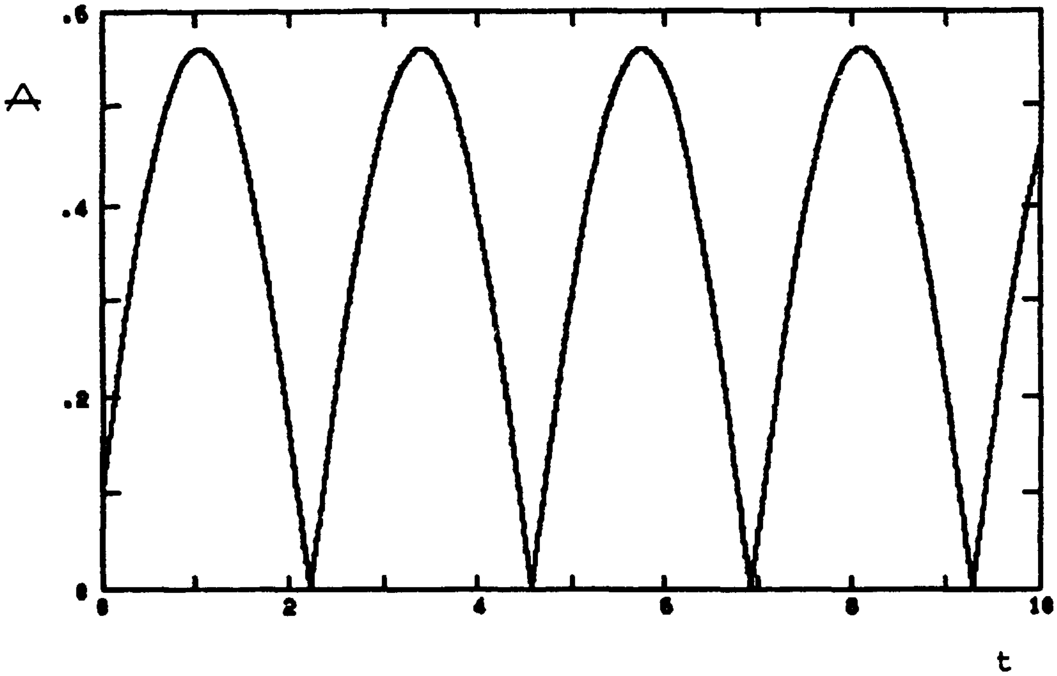
This has the integral

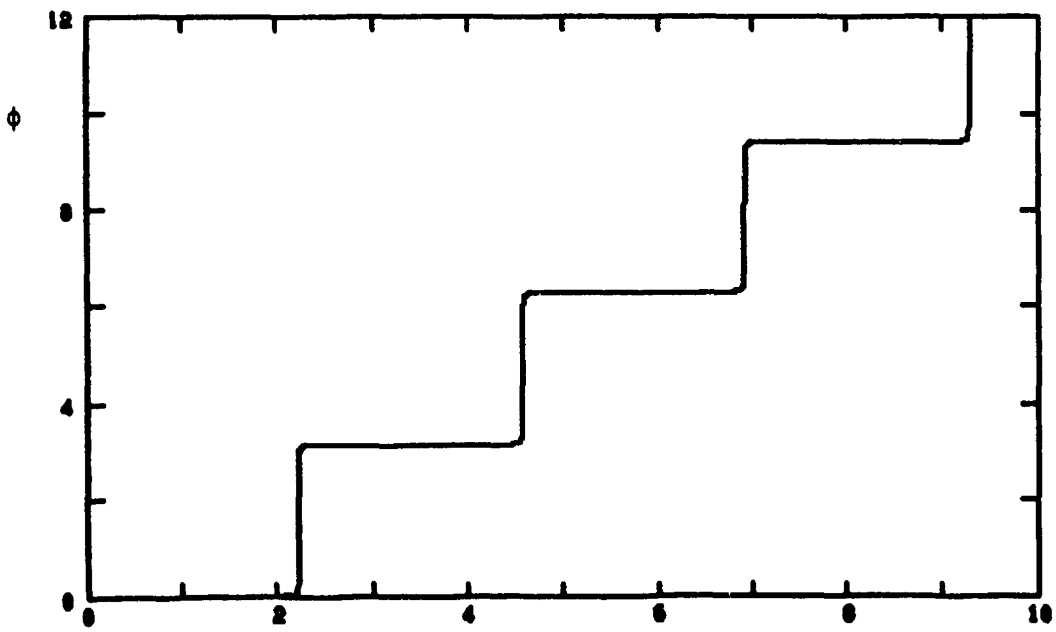
$$\frac{1}{2}\dot{\mathcal{A}}^2 + V(\mathcal{A}) = \mathcal{E}$$

where

$$V = \frac{1}{2}\left(b^2/\mathcal{A}^2 - \lambda k^4 \mathcal{A}^2 - \frac{3}{8}\nu k^5 \mathcal{A}^4\right)$$

and $\mathcal{E}$ is a constant. Solutions may be expressed in elliptic functions, but it is instructive simply to look at plots of the amplitude and phase, here for b = .001, λ = -2 and k = 1.





THE CAVEATS

We have suggested that a certain kind of variability is caused by instabilities in thin layers on large horizontal scales with long periods. This opens the way for an analysis such as we have just sketched. The calculations may be elaborate, but they are feasible. Their astrophysical interest lies in the relative ease with which they may be extended to allow for nonBoussinesq effects [25,26], when those are not too pronounced. Thus one can treat compressibility, κ-mechanism, and so on, and the extension to more general boundary conditions has been studied [27]. In other words, a number of features of nonlinear nonradial pulsation can be studied along the lines we have outlined here. It only requires finding the right instability.

However, the main outcome so far is only qualitative because (18) has an infinite number of possible solutions and we have not given a method for selecting one from among them. The removal of this degeneracy requires the introduction of higher order information and this may be effected by procedures that we shall go into elsewhere. Nevertheless, the cyclic character of the solutions is a correct, if particular, asymptotic consequence of the equations.

Yet equation (18) is too tame and it gives only periodic solutions. Nor do the higher order corrections remove this failing. But we do have a situation with two competing instabilities such as the models mentioned at the beginning did. If they give chaotic solutions [5,28] why do we not find them from (18)? The formal difference is simply that (18) is second order whereas the model equations are third order. If we analyze the latter in the neighborhood of the onset of the instabilities, we may reduce them to second order equations. Chaos in the models arises only for highly unstable conditions and was discovered by numerical means. If we want to study strongly nonlinear conditions in the problems discussed here, we have to solve nonlinear partial differential equations, and that is a far more difficult task than solving the model equations, which are ordinary differential equations. Perhaps when the full numerical solutions are found, there will appear just the rich structure we see in the solar cycle; that is certainly one thing that should be attempted. But there is another way to enrich the time dependence of our model that we believe is relevant to the solar case and must be included in any event.

If we consider a case with three competing instabilities, the procedures described here lead to a third order amplitude equation near a critical point at which all three instabilities begin at once. We have been studying these questions with colleagues in Nice and that work [29] will provide some concrete examples of what we mean. For our present purposes we need only the simple extension of the idea of competing instabilities to the case of three instablities. It turns out that in each of the examples of variable variability cited here, there seem to be three competing instabilities (at least potentially) in the models that are being proposed. Consider the case of solar variability and the picture that it is driven from the base of the convective zone [13].

In the process of solar spin down [30], it has been suggested, when the hydromagnetic torques of the solar wind brake the convective zone, it in turn pumps a secondary flow into the subconvective layers [31]. The convective pumping process is resisted by the stable layers, hence it can penetrate only into a shallow layer below the convective zone [30,32]. This layer initially supports the rotational difference between the convective zone and the radiative interior, but it ultimately loses stability. The further developments are not fully understood, but one of the plausible possibilities is that the resulting motions maintain the layer in conditions near to marginal instability. This is the source of the third competing instability that we believe must be included in the description of the solar activity cycle. For schematic versions of this problem, if the geometry is right, one gets third order equations for the amplitude of the motion. Therein, we suggest, lies the cause of some of the chaos of the solar cycle.

This hint of further developments only underscores how incomplete is the picture we have presented here. But at least we have been able to see one direction to go in which to find the source of variable variability. The temporal behaviors that are emerging at this stage of the work have some of the right kind of qualitative behavior and the mechanism of competing instabilities seems to provide a possible basis for understanding the examples of stellar variability that we have mentioned here.

REFERENCES

1. Eddy, J. A.: Historical and Arboreal Evidence for a Changing Sun, The New Solar Physics, ed. J. A. Eddy, 1978, Westview Press, Boulder, pp. 11-34.

2. Robinson, E. L.: The Observational Properties of the ZZ Ceti Stars, IAU Colloquium No. 53, White Dwarfs and Variable Degenerate Stars, ed. H. M. Van Horn and V. Weicheman, 1979, Univ. of Rochester, Rochester, pp. 343-357.

3. Scherrer, P. H.; Wilcox, J. M.; Kotov, V. A.; Severny, A. B.; and Tsap, T. T.: Observations of Solar Oscillations with Periods of 160, Nature, vol. 277, 1979, pp. 635-637.

4. Marsden, J. E. and McCracken, M.: The Hopf Bifurcation and its Applications, 1976, Springer-Verlag, New York.